# REAL-WORLD OUTCOMES OF IMMUNOTHERAPY IN LUNG CANCER PATIENTS

## PROMPT EVIDENCE GENERATION BEYOND CLINICAL TRIALS

MARJON VERSCHUEREN

# Real-world outcomes of immunotherapy in lung cancer patients

## Prompt evidence generation beyond clinical trials

**Marjon Verschueren**

**Colofon**
© M.V.Verschueren, Amsterdam, The Netherlands, 2026
ISBN: 978 94 6537 565 6

Provided by thesis specialist Ridderprint, ridderprint.nl
Printing: Ridderprint
Layout and design: Tara Schollema, persoonlijkproefschrift.nl
Cover design: Gayle Tjong-Kim-Sang, persoonlijkproefschrift.nl

Printing of this thesis was financially supported by Utrecht Institute of Pharmaceutical Sciences (UIPS), de eenheid Farmacie van het St Antonius ziekenhuis Nieuwegein/Utrecht.

# Real-world outcomes of immunotherapy in lung cancer patients

## Prompt evidence generation beyond clinical trials

Uitkomsten van immunotherapie bij longkankerpatiënten in de klinische praktijk
Prompt bewijs genereren in aanvulling op klinische studies
*(met een samenvatting in het Nederlands)*

**Proefschrift**

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht op gezag van de
rector magnificus, prof. dr. ir. W. Hazeleger,
ingevolge het besluit van het College voor Promoties
in het openbaar te verdedigen op
maandag 8 juni 2026 des middags te 2.15 uur.

door

**Marjon Valentina Verschueren**
geboren op 13 februari 1991
te Nijmegen

**Promotoren:**
Prof. dr. E.M.W. van de Garde
Prof. dr. A.C.G. Egberts

**Copromotoren:**
Dr. B.J.M. Peters
Dr. L.T. Bloem

**Beoordelingscommissie:**
Dr. V.H.M. Deneer
Prof. dr. H. Gardarsdottir
Prof. dr. M.M. van den Heuvel
Prof. dr. A.D.R. Huitema
Prof. dr. M.W.J.M. Wouters

# Table of contents

# CHAPTER 1

## General introduction

## Treating patients with advanced non-small cell lung cancer without targetable driver mutations

Lung cancer is the most common type of cancer worldwide, with approximately 2.5 million newly diagnosed patients each year, accounting for 12.4% of new cancer cases globally. [1] Non-small cell lung cancer (NSCLC) is the most prevalent form, representing 85% of all lung cancer cases. Unfortunately, in about 50% of patients, NSCLC is diagnosed at an advanced stage, which is associated with a poor prognosis. Historically, these patients have had 5-year survival rates of approximately 5%. [2]

For many decades, the standard of care for patients with advanced NSCLC without targetable mutations was chemotherapy, which offered limited survival benefits and was associated with significant toxicity. However, the treatment landscape for these patients has rapidly evolved since the introduction of immune checkpoint inhibitors (ICIs). ICIs are monoclonal antibodies that target immune regulatory pathways, such as the CTLA-4, PD-1, or PD-L1 pathways. By blocking these inhibitory signals, ICIs enhance the immune system's ability to recognize and eliminate cancer cells. [3, 4]

For patients with advanced NSCLC, nivolumab (anti-PD-1) was the first ICI approved by the European Medicines Agency in 2015, followed by pembrolizumab (anti-PD-1) in 2016, atezolizumab (anti-PD-L1) in 2017, durvalumab (anti-PD-L1) in 2018, cemiplimab (anti-PD-1) in 2019, and ipilimumab (anti-CTLA-4) in 2020. [5-10] Initially, these treatments were approved for use as monotherapies, but their indications have expanded to include combinations with chemotherapy (e.g., pembrolizumab plus chemotherapy) and other ICIs (e.g., ipilimumab plus nivolumab). [11,12] As a result, the standard of care for patients with advanced NSCLC without targetable mutations has shifted from chemotherapy to immunotherapy-centered treatment strategies.

## Complementing clinical trial outcomes with real-world outcomes

As for all medicines, the lifecycle of a new ICI begins with preclinical studies, followed by phase I and II trials that establish safety, dosing, and preliminary efficacy. Phase III RCTs then evaluate efficacy and safety under controlled conditions and against the current standard of care, providing the primary evidence base for regulatory approval and further policy-related and clinical decision-making. Given that these RCTs include highly selected patient populations and treatment adheres to standardized treatment and follow-up protocols, RCT outcomes may have limited generalizability to the broader and more heterogeneous patient population in clinical practice. [13,14] Several studies have shown that survival outcomes observed in clinical practice are less favorable than those observed in clinical trials. For example, Cramer-van der Welle et al. reported that the overall survival (OS) for patients with advanced NSCLC treated with first-line chemotherapy or targeted therapy was nearly 25% shorter in clinical practice than in the corresponding RCTs. [15]

There are several reasons why outcomes differ between RCTs and clinical practice. First, patients treated in clinical practice are usually older, less fit, and have more comorbidities than those treated in RCTs. Second, treatment adherence is usually lower in clinical practice than in RCTs. [15] Third, monitoring and follow-up are less standardized, with less frequent and consistent response assessments in clinical practice than in RCTs what could cause biased times to outcomes.[16] The limited generalizability of RCT outcomes has important implications such as less informed clinical decision making because of lack of knowledge about effects in patients outside trial eligibility criteria [17] Generating real-world evidence (RWE) helps to overcome this knowledge gap. RWE is generated through observational studies that use real-world data (RWD), which refers to routinely collected information about patients' health status and the delivery of healthcare. [18] Such data come from sources outside traditional RCTs, including electronic health records (EHRs), insurance claims, patient registries, and digital health tools such as wearable devices. [18] RWD complements RCT data by providing insights into treatment outcomes in more heterogeneous patient populations and insight into treatment outcomes over more extended follow-up periods. In addition, RWD allows head-to-head comparisons of treatments that have not been, and probably never will be, compared in RCTs. [19,20] The use of RWE is increasingly considered valuable for regulatory and clinical decision-making. In regulatory decision-making, RWE is currently still primarily used for post-approval safety monitoring, but its potential for evaluating treatment effectiveness is increasingly being recognized [18,19]. In clinical decision-making, RWE is becoming increasingly relevant for guiding individual treatment decisions for patients underrepresented in RCTs. [22, 23]

## Methodological challenges for prompt evidence generation

Individual treatment decisions are becoming increasingly complex due to the diversity of patient characteristics in clinical practice and the rapidly expanding range of treatment options. In this dynamic setting, prompt insights into their real-world effectiveness and safety are needed to ensure that patients receive the most appropriate treatment. Prompt generation of RWE can help clinicians to quickly identify patients who benefit most from new treatment options and detect ineffective or harmful treatments. [24] Ideally, real-world effectiveness and safety should be monitored continuously and compared with those of other treatments to enable timely adjustments to treatment strategies. However, generating such prompt insights based on observational studies is challenging. Manual review of EHRs (still considered the gold standard for data collection) is unsuitable for continuous monitoring, as it would require repeated reviews of patient records to capture information on treatment outcomes (near) real-time. Another barrier is the lack of standardization of data collection and analyses performed for observational studies, which is essential not only for prompt evidence generation but also for ensuring valid findings across studies. In addition, repeated analyses during the maturation of data using conventional statistical methods could increase the risk of false-positive findings due to multiple testing. These challenges highlight the need for methodological advancements to support the prompt and reliable generation of RWE, enabling treatment decisions based on the most up-to-date information.

## Objective of this thesis

The overall objective of this thesis is to complement survival outcomes data reported in RCTs with data from patients with advanced NSCLC treated with immunotherapy in clinical practice, while also evaluating methodological challenges and advancements for prompt evidence generation. The resulting evidence is intended to support patients and clinicians in making well-informed treatment decisions.

## Outline of this thesis

**Table 1** summarizes the contents of this thesis based on the study design aspects. **Chapter 2** evaluates the outcomes of patients with advanced NSCLC treated with immunotherapy in clinical practice. **Chapter 2.1** compares the real-world survival outcomes of patients with locally advanced NSCLC treated with adjuvant durvalumab with those reported in the PACIFIC RCT. **Chapter 2.2** examines the real-world survival outcomes of pembrolizumab plus chemotherapy in advanced NSCLC, stratified by PD-L1 expression and histology, and compares these results with the KEYNOTE-189 and KEYNOTE-407 RCTs.

**Chapter 2.3** compares real-world treatment responses and progression-free survival (PFS) between nivolumab–ipilimumab plus chemotherapy and pembrolizumab plus chemotherapy in PD-L1–negative patients with advanced NSCLC. **Chapter 2.4** investigates the impact of concomitant medication on the microbiome and its effect on real-world survival outcomes in advanced NSCLC patients treated with immunotherapy, using a historical chemotherapy-matched cohort design.

**Chapter 3** addresses methodological challenges and advancements for prompt evidence generation. **Chapter 3.1** explores how PFS is defined, measured, and reported in RCTs and observational studies of patients with metastatic NSCLC treated with immunotherapy. **Chapter 3.2** develops and evaluates a text-mining algorithm for prospectively capturing disease progression in EHR data from patients with metastatic NSCLC receiving immunochemotherapy. **Chapter 3.3** presents a Bayesian survival model that continuously compares accumulating survival data from clinical practice with fixed RCT data to assess whether survival outcomes differ, thereby enabling prompt and interpretable insights to inform clinical and policy-related decision-making.

Finally, **Chapter 4** discusses the results of the preceding chapters and their context and provides recommendations for the future.

**Table 1.** Overview of studies presented in this thesis by study design aspects

| | Patient population | Setting | Exposure/intervention | Comparator(s) | Outcome(s) |
|---|---|---|---|---|---|
| **Chapter 2:** Outcomes of immunotherapy in lung cancer patients treated in clinical practice | | | | | |
| **Chapter 2.1** | Locally advanced NSCLC | Clinical practice (EHR) | CRT with adjuvant durvalumab | 1. CRT alone (clinical practice)<br>2. RCT (PACIFIC) | PFS, OS |
| **Chapter 2.2** | 1. Non-squamous metastatic NSCLC<br>2. Squamous metastatic NSCLC | Clinical practice (EHR) | Pembrolizumab + chemotherapy (first-line) | 1. RCT (KEYNOTE-189)<br>2. RCT (KEYNOTE-407) | PFS, OS by PD-L1 stratum |
| **Chapter 2.3** | PD-L1 negative metastatic NSCLC | Clinical practice (EHR) | Nivolumab-ipilimumab + chemotherapy (first-line) | Pembrolizumab + chemotherapy (first-line) | ORR, PFS |
| **Chapter 2.4** | Metastatic NSCLC treated with an ICI (matched with historical controls treated with chemotherapy) | Clinical practice (EHR) | Comedication | No comedication | OS |
| **Chapter 3:** Methodological challenges and advancements for prompt evidence generation | | | | | |
| **Chapter 3.1** | Metastatic NSCLC, treated with an ICI | NA (published literature) | Observational studies | RCTs | Definition, measurement and reporting of PFS |
| **Chapter 3.2** | Metastatic NSCLC, treated with pembrolizumab + chemotherapy | Clinical practice (EHR) | Text-mining algorithm | Manual data collection | PFS |
| **Chapter 3.3** | Metastatic NSCLC, treated with pembrolizumab with or without chemotherapy | Clinical practice (EHR) + RCT | Bayesian survival modelling | Cox proportional hazards regression modelling | OS |

**Abbreviations:** CRT, chemoradiotherapy; EHR, electronic health records; ICI, immune checkpoint inhibitor; NA, not available; NSCLC, non-small-cell lung cancer; ORR, Objective response rate; OS, Overall survival; PD-L1, programmed death-ligand 1;PFS, progression free survival;RCT, Randomized controlled trial

## References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2022: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2023;73(1):17-48
2. Ganti AK, Klein AB, Cotarla I, Seal B, Chou E. Update of Incidence, Prevalence, Survival, and Initial Treatment in Patients With Non–Small Cell Lung Cancer in the US. *JAMA Oncol.* 2021;7(12):1824–1832.
3. Herzberg B, Campo MJ, Gainor JF. Immune checkpoint inhibitors in non-small cell lung cancer. Oncol. 2016; 22(1):81-88.
4. Doroshow DB, Sanmamed MF, Hastings K, Politi K, Rimm DL, Chen L, et al. Immunotherapy in non-small cell lung cancer: facts and hopes. *Clin Cancer Res*. 2019;25(15):4592-4602
5. European Medicines Agency (EMA). Opdivo: European Public Assessment Report (EPAR). 2015. EMA/120822/2025
6. European Medicines Agency (EMA). Keytruda: European Public Assessment Report (EPAR). 2016. EMA/99022/2025
7. European Medicines Agency (EMA). *Tecentriq: European Public Assessment Report* (EPAR). 2017. EMA/361473/2024.
8. European Medicines Agency (EMA). Imfinzi: European Public Assessment Report (EPAR). 2018. EMA/196154/2025.
9. European Medicines Agency (EMA). *Libtayo: European Public Assessment Report* (EPAR). 2019. EMA/101482/2023.
10. European Medicines Agency (EMA). *Yervoy: European Public Assessment Report* (EPAR). 2011. EMA/85145/2025
11. Gandhi L, Rodriguez-Abreu D, Gadgeel S, Esteban E, Felip E, De Angelis F, et al. Pembrolizumab plus chemotherapy in Metastatic Non-Small-Cell Lung Cancer. N Engl J M. 2018; 378(22) : 2079-2092.
12. Paz-Ares L, Cioleanu T, Cobo M, Schenker M, Zurawski B, Menezes J, et al. First-line nivolumab plus ipilimumab combined with two cycles of chemotherapy in patients with non-small-cell lung cancer (CheckMate 9LA): an international randomised, open-label, phase 3 trial. Lancet Oncol. 2021;22(2):198-211.
13. Wilson BE, Hanna TP, Booth CM. Efficacy-effectiveness gaps in oncology: Looking beyond survival. Cancer.2024;130(3):335-338.
14. Gupta A, Nguyen P, Wilson BE, Booth CM, Hanna TP. Health care contact days and outcome in clinical trials vs routine care among patients with non-small-cell lung cancer. JAMA Netw Open.2025;8(4):e255033.
15. Cramer-van der Welle CM, Peters BJM, Schramel FMNH, Klungel OH, Groen HJM, van de Garde EMW. Real-world outcomes versus clinical trial results of systemic treatments in metastatic non-small-cell lung cancer in the Netherlands. *Eur Respir J.* 2018;52(6):1801100.
16. Nordon C, Karcher H, Groenwold RHH, Ankarfeldt MZ, Pichler F, Chevrou-Severac H, et al. The “efficacy-effectiveness gap”: Historical background and current conceptualization. Value in Health. 2016;19(1):75-81.
17. Brinkhuis F, Goettsch WG, Mantel-Teeuwisse AK, Bloem LT. High cost oncology drugs without proof of added benefit are burdening health systems. *BMJ*. 2024;384:q511
18. European Medicines Agency. Real-World-Evidence (RWE). Amsterdam: European Medicines Agency; 2024. EMA/152628/2024
19. Azoulay L. Rationale, strengths, and limitations of real-world evidence in oncology: a Canadian review and perspective. Oncologist. 2022; 27(9):e731-738.

20. Berger ML, Sox H, Willke RJ, Brixner DL, Eichler HG, Goettsch W, et al. Good practices for real-world data studies of treatment and/or comparative effectiveness: recommendations from the joint ISPOR-ISPE special task force on real-world evidence in health care decision making. Pharmacoepidemiol Drug Saf. 2017; 26(9): 1033-1039
21. Hennessy S, Atsuta Y, Hill S, Rägo L, Juhaeri J. Real-world data and real-world evidence in regulatory decision making: Report summary from the council for international organizations of medical sciences (CIOMS) working group XIII. Pharmacoepidemiol Drug Saf. 2025; 34(3):e70117
22. Costa V, Graziano Custodio M, Gefen E, Fregni F. The relevance of the real-world evidence in research, clinical, and regulatory decision making. Front Public Health. 2025;13:1512429
23. Miksad RA and Calip GS. Future of cancer treatment guidelines: Integrating real-world insights for equitable cancer care. JCO Clin Cancer Inform. 2024; 8:e2400081.
24. Eichler H, Bloechl-Daum B, Broich K, Kyrle PA, Oderkirk J, Rasi G. Data rich, information poor: Can we use electronic health records to create a learning healthcare system for pharmaceuticals? Clin Pharmacol Ther. 2019; 105(4):912-922.

# CHAPTER 2

# Outcomes of immunotherapy in lung cancer patients treated in clinical practice

# 2.1

# **Durvalumab after chemoradiotherapy in patients with stage III NSCLC:**
## Real-world outcomes versus clinical trial results

Marjon V Verschueren, Talitha Dijs, Judith L Gulikers, Ard van Veelen,
Sander Croes, Lizza EL Hendriks, Adrianus AJ Smit, Lourens T Bloem,
Antoine CG Egberts, Ewoudt MW van de Garde, Bas JM Peters

*Immunotherapy, 2023 Aug; 15 (11): 839-851*
*DOI: 10.2217/imt-2023-0002.*

# Abstract

## Objective

We investigated the effectiveness of durvalumab post-concurrent CRT (cCRT) and post-sequential CRT (sCRT) versus cCRT and sCRT alone and compared these outcomes with the PACIFIC trial.

## Methods

Four cohorts of stage III NSCLC patients who received CRT were included: cCRT with and without durvalumab, sCRT with and without durvalumab. PFS and OS were analyzed using Cox regression.

## Results

Durvalumab improved PFS (cCRT: aHR=0.74, sCRT: aHR=0.71) and OS (cCRT: aHR=0.71, sCRT: aHR=0.32) , although not all results were significant. PFS was longer in the real-world than in the trial, while OS did not differ.

## Conclusion

Durvalumab after CRT improved the survival outcomes. The difference between PFS in our study and the trial may be due to differences in follow-up methods.

## Introduction

Approximately 20-35% of patients with Non-Small Cell Lung Cancer (NSCLC) are diagnosed with locally advanced disease [1]. For decades, the standard treatment for these patients was concurrent or sequential chemoradiotherapy (cCRT or sCRT) [2]. However, this standard treatment regime changed with the results of the PACIFIC trial demonstrating a progression free survival (PFS) and overall survival (OS) benefit for durvalumab versus placebo after cCRT [3,4]. As a result, durvalumab after chemoradiotherapy (CRT) has become the first Immune checkpoint inhibitor approved for treating stage III NSCLC patients, and is currently considered the standard of care [5,6].

Nevertheless, treatment effects observed in randomized clinical trials (RCT) may not be observed in daily clinical practice [7,8], since these patients are usually older, have more comorbidities and worse performance states which could lead to worse baseline prognosis. Additionally, follow up of patients treated in daily clinical practice is less standardized than in clinical trials. Furthermore, in clinical practice, durvalumab after sCRT is usually offered to less fit patients [5,6], which may further influence their survival outcomes given that sCRT is also less effective compared to cCRT. [9]Real-world studies have the potential to address these knowledge gaps.

To date, 21 real-world studies evaluated the effectiveness of durvalumab in stage III NSCLC patients [10-30] . A meta-analysis summarized 13 of these studies (n = 1885) and reported marginally better OS survival rates than reported in the PACIFIC trial (12-month OS: 90% vs. 83.3%) [31]. Also the large single arm, observational PACIFIC-R study (n = 1399) suggested that the effectiveness of durvalumab in the real-world is at least comparable with the clinical trial [10]. However, recently another large study by Sanker et.al (2022) demonstrated that the OS of durvalumab in the real-world is shorter than in the trial (median OS: 34.7 vs. 47.5 months), with an effectiveness-efficacy gap (EE gap) of 0.72 [26]. So far, this was the only study that quantified the difference between the real-world study and PACIFIC trial (expressed as EE gap) whereas other studies (n=20) only provided a descriptive comparison. Moreover, most observational studies (n=12) did not assess the relative effectiveness of durvalumab because they lack a control group [10,13-15, 17–22, 27,28]. Lastly, only five studies included patients treated with sCRT [10,11,13,22,27].

Here, we assess the real-world effectiveness of durvalumab in a Dutch cohort of stage III NSCLC patients who received durvalumab after cCRT or sCRT versus stage III NSCLC patients who received cCRT or sCRT alone. Additionally, we compare the real-world survival outcomes to the outcomes of the PACIFIC trial by reconstructing individual patient data of the trial.

## Methods

### Setting, design and study population

We conducted a multicenter, retrospective, cohort study in five hospitals spread out geographically over the Netherlands, including four large teaching hospitals (St Antonius Hospital Utrecht/Nieuwegein, Canisius Wilhelmina Hospital Nijmegen, Catharina Hospital Eindhoven, OLVG Amsterdam) and one academic centre (Maastricht University Medical Center+). Patients diagnosed with stage III NSCLC and treated with CRT between January 2012 and December 2021 were identified in these hospitals. The following methods were used to select eligible patients; the OLVG used the Clinical Data Collector (CTcue®), the MUMC used their local developed and Gamp5 validated software application for data collection, and the SAZ, CWZ and CZE used the database of the Netherlands Cancer Registry.

Patients were excluded if they; received less than two cycles of platinum-based chemotherapy during CRT; had disease progression before the completion of CRT; or had lung resection during the time that CRT was applied. The population was divided into the following four cohorts; 1) cCRT with durvalumab, 2) cCRT without durvalumab, 3) sCRT with durvalumab and 4) sCRT without durvalumab. Durvalumab patients received at least one dose of adjuvant durvalumab. Patients treated without durvalumab were referred to as historical controls because they were treated with CRT alone in the pre-durvalumab era (before 1st of april 2018). Patients treated with CRT alone in the durvalumab-era were excluded.

Baseline characteristics were extracted from the patients' Electronic Health Records between 30 days before and 30 days after the date of diagnosis. Follow-up information was collected until the end of this study on 1 July 2022. The study design is schematically depicted in Figure 1.

Study data were collected, anonymized, and managed using REDCap electronic data capture tools [32]. All methods were carried out in accordance with relevant guidelines and regulations. The ethics committees, the Santeon Institutional Review Board (SBD 2021-001) and the academic hospital Maastricht/University Maastricht (2021-2843), approved the study and waived the need for informed consent. The study was performed in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

## Outcomes

The primary outcome PFS, expressed in months, was calculated from the index-date till the date of disease progression or death, or till the date of censoring, whichever occurred first. The date of disease progression was determined by the earliest medical note from the thoracic oncologist stating that the disease had progressed. The secondary outcome OS, expressed in months, was calculated from the index-date till the date of death or censoring. Patients without progression and/or those who were still alive at the end of follow-up were censored at the date of their last clinic visit.

The start date of durvalumab treatment was used as index date for patients treated with durvalumab. Because historical controls did not receive durvalumab, imputed index dates were used which were calculated in two steps. First, the median time between the start date of cCRT or sCRT and the start date of durvalumab was calculated. Second, for both the historical controls treated with cCRT and sCRT these median times were added on to the start date of CRT to generate an imputed index date. Patients within the historical control cohort who progressed, died or were censored before the imputed index date were excluded (n = 23) to avoid immortal time bias [33]. The median time between the start of sCRT and start of durvalumab was 146 days [range 97 – 188 days] and the median time between start of cCRT and durvalumab was 91 days [range 49 – 189 days] (Figure 1).

## Potential confouders and/or effect modifiers

The following characteristics were included in the analyses: age, gender, body mass index (BMI), Eastern Cooperative Oncology Group-Performance status (ECOG-PS), histology subtype and disease stage. All patient characteristics were extracted within 30 days before or after diagnosis (Figure 1).

## Data analysis

Statistical software (R version 4.1.2) was used to conduct the data analyses. Descriptive statistics for categorical variables were reported as the number of observations (proportions), while the mean [± standard deviation (SD)] was provided for normally distributed continuous data and the median [range] was provided for non-normally distributed continuous data. Chi-square and t-tests were used to compare the baseline characteristics between patients treated with and without durvalumab.

The Kaplan-Meier (KM) method was used to visualize survival curves for PFS and OS for the four cohorts. For both the sCRT and cCRT regime, patients treated with durvalumab were compared to patients treated without durvalumab (historical controls). For this comparison simple and multiple Cox regression were used to calculate unadjusted and adjusted Hazard Ratios (HR) and their 95% Confidence Intervals (CIs)). The covariates mentioned above were tested with simple Cox regression analysis for both cohorts to identify variables associated with PFS or OS. All variables with a p-value ≤ 0.20 and three other variables; age, gender and the type of treatment (durvalumab or historical control, were used to construct multiple Cox regression models. Multiple imputation (MI) was used to impute the missing observations for ECOG PS under the assumption that data were missing at random. Lastly, for both cCRT patients treated with durvalumab and without durvalumab we compared the effectiveness outcomes of our cohort (the 'real-world') to efficacy outcomes of the PACIFIC cohort by estimating HRs.Therefore, the algorithm developed by Guyot and colleagues was used to reconstruct individual patient data from the published OS and PFS curves from the PACIFIC trial [3, 34, 35)].

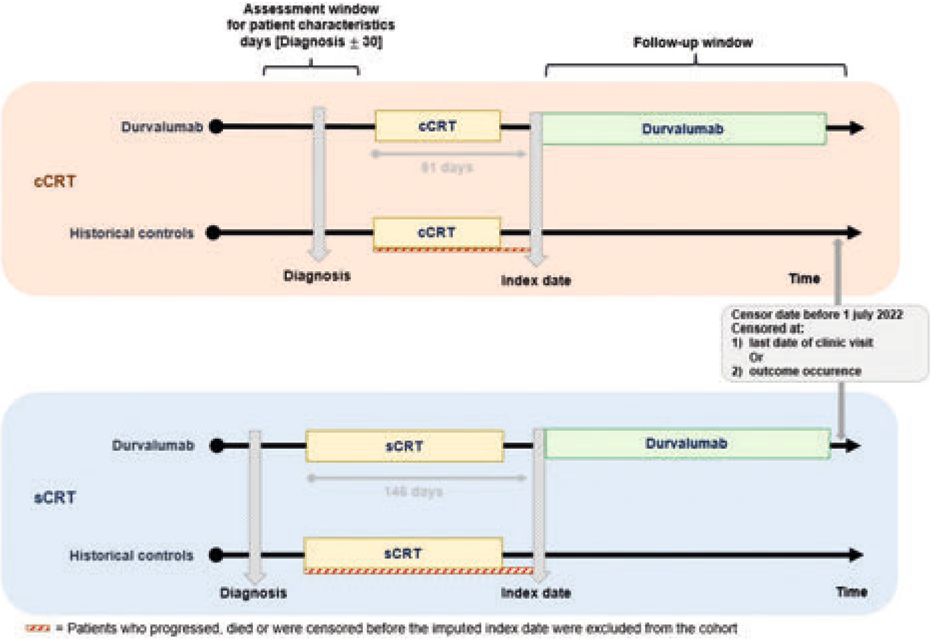


**Figure 1.** Study overview.

Legend: The orange rectangle displays the timeline for patients treated with concurrent chemoradiotherapy (cCRT) with durvalumab or without durvalumab (historical controls). The blue rectangle displays a timeline for patients treated with sequential chemoradiotherapy (sCRT) with durvalumab or without durvalumab (historical controls).

## Results

### Baseline characteristics

We included 267 cCRT patients, of which 106 were treated with durvalumab and 161 were treated without durvalumab (controls), and 116 sCRT patients, of which 21 were treated with durvalumab and 95 without durvalumab (controls). For the cCRT group, patients treated with durvalumab had a higher BMI (26.0 versus 24.8 kg/m2), more often an ECOG PS 1 (38.7% versus 30.4%), and less often a disease stage IIIa (41.5% versus 51.6%) compared with the historical controls (Table 1). In the sCRT group, patients treated with durvalumab were younger (64.5 versus 69.2 year) and had less often disease stage IIIb (38.1% versus 55.8%) compared with the historical controls.

Medians for the time between the end of CRT and the start of durvalumab, number of durvalumab administrations per patient and treatment duration can be found in Table 1. The median follow-up time was 15.4 (range: 1.0 – 46.1) months for patients treated with durvalumab and 25.3 (range: 0.5 – 92.5) months for patients treated without durvalumab.

**Table 1.** Patient characteristics of stage III NSCLC patients treated with cCRT and sCRT with or without durvalumab (historical controls).

| **Characteristics** | | **cCRT (n = 267)** | | | **sCRT (n = 116)** | | |
|---|---|---|---|---|---|---|---|
| | | **Durvalumab (n = 106)** | **Historical controls (n = 161)** | **p-value** | **Durvalumab (n=21)** | **Historical controls (n = 95)** | **p-value** |
| Age, years (mean, sd) | | 64.2 (9.5) | 63.4 (9.3) | 0.51 | 64.5 (9.7) | 69.2 (8.2) | 0.02 |
| Sex (male, n (%)) | | 54 (50.9) | 86 (53.4) | 0.84 | 8 (38.1) | 54 (56.8) | 0.19 |
| BMI, kg/m$^2$ (mean, sd) | | 26.0 (4.5) | 24.8 (4.5) | 0.04 | 25.9 (3.1) | 24.5 (3.9) | 0.08 |
| ECOG-PS (n (%)) | | | | 0.05 | | | 0.12 |
| | 0 | 62 (58.5) | 90 (55.9) | | 9 (42.9) | 39 (41.1) | |
| | 1 | 41 (38.7) | 49 (30.4) | | 11 (52.4) | 44 (46.3) | |
| | ≤2 | 1 (0.9) | 6 (3.7) | | 1 (4.8) | 7 (7.4) | |
| | unknown | 2 (1.9) | 16 (9.9) | | 0 | 5 (5.3) | |
| Disease stage (n, (%)) | | | | 0.03 | | | <0.01 |
| | IIIa | 44 (41.5) | 83 (51.6) | | 7 (33.3) | 35 (36.8) | |
| | IIIb | 46 (43.4) | 70 (43.5) | | 8 (38.1) | 53 (55.8) | |
| | IIIc | 6 (5.7) | 3 (1.9) | | 4 (19.0) | 7 (7.4) | |
| | unknown | 10 (9.4) | 5 (3.1) | | 2 (9.5) | 0 | |
| Histology (n, (%)) | | | | 0.39 | | | 1.0 |
| | Squamous | 32 (30.2) | 58 (36.0) | | 8 (38.1) | 35 (36.8) | |
| | Nonsquamous | 74 (69.8) | 103 (64.0) | | 13 (61.9) | 60 (63.2) | |
| Days between end radiation and start durvalumab (median[range]) | | 41 [6-124] | - | | 47 [8-93] | - | |
| Number of durvalumab administrations (median[range]) | | 16 [1-27] | - | | 10 [1-27] | - | |
| Months of treatment duration durvalumab (median[range]) | | 8 [1-40] | - | | 5 [1-13] | - | |

Abbreviations: cCRT, concurrent chemoradiotherapy; sCRT, sequential chemoradiotherapy; BMI, Body Mass Index; ECOG-PS, Eastern Cooperative Group performance status.

## Survival outcomes

### *Progression-free survival*

In the cCRT cohorts, the observed median PFS (mPFS) for durvalumab patients was longer (27.1 months [ 95%CI, 19.2 – not reached (NR)] versus 15.0 months [95%CI, 12.3 – 20.7]) and the 12-months PFS rate was higher (70.2% versus 59.0%) than for historical control patients (Figure 2). The unadjusted HR 0.68 (95%CI, 0.48 – 0.97) and adjusted HR 0.69 (95%CI 0.49 – 0.99) for PFS were better for patients treated with durvalumab than for patients treated without durvalumab. None of the other variables were significantly associated with PFS in the multiple Cox regression model (Table 2).

In the sCRT cohorts, the observed mPFS for durvalumab patients was slightly longer (8.0 months [95%CI, 4.6 – NR) versus 7.4 months [95%CI, 6.1 – 10.3]) and the 12-months PFS rate was higher (49.2% versus 32.6%) than for historical control patients. The unadjusted HR 0.77 (95%CI 0.42 – 1.42)) and adjusted HR 0.71 (95%CI 0.37 – 1.40) for PFS were better for patients treated with durvalumab than for patient treated without durvalumab, although these results were not significant. None of the other variables were significantly associated with PFS in the multiple Cox regression model (Table 2).

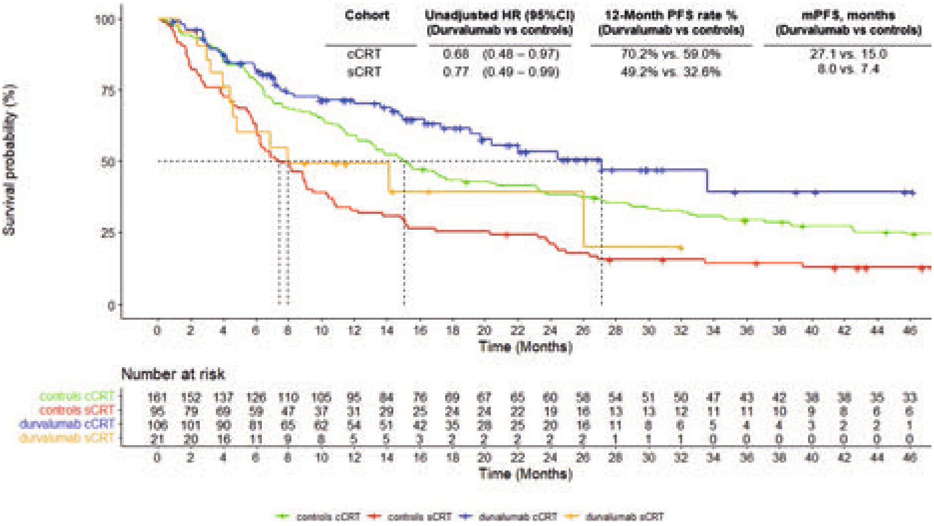


**Figure 2.** Kaplan-Meier curves for progression-free survival for patients who received cCRT or sCRT with and without durvalumab.

**Table 2.** Progression free survival. Simple- and multiple cox-regression model for the cCRT and sCRT population.

| | cCRT (n = 267) | | | | | | sCRT (n = 116) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Simple model** | | | **Multiple model** | | | **Simple model** | | | **Multiple model** | | |
| | *HR* | *95% CI* | *p-value* | *HR* | *95% CI* | *p-value* | *HR* | *95% CI* | *p-value* | *HR* | *95% CI* | *p-value* |
| Type of treatment | | | | | | | | | | | | |
| Control | Ref. | 0.48 – 0.97 | 0.04 | 0.69 | 0.49 – 0.99 | 0.04 | Ref. | 0.42-1.41 | 0.40 | 0.71 | 0.37 – 1.40 | 0.33 |
| Durvalumab | 0.68 | | | | | | 0.77 | | | | | |
| Age | | | | | | | | | | | | |
| <75 years | Ref. | 0.71 – 1.77 | 0.64 | 1.17 | 0.73 – 1.89 | 0.51 | Ref. | 0.48-1.19 | 0.23 | 0.80 | 0.49 - 1.30 | 0.36 |
| ≥ 75 years | 1.12 | | | | | | 0.76 | | | | | |
| Gender | | | | | | | | | | | | |
| Male | Ref. | 0.75 – 1.37 | 0.94 | 1.02 | 0.75 – 1.39 | 0.91 | Ref. | 0.67-1.37 | 0.82 | 0.70 | 0.49 - 1.30 | 0.11 |
| Female | 1.01 | | | | | | 0.96 | | | | | |
| BMI | | | | | | | | | | | | |
| <30 | Ref. | 0.53 – 1.27 | 0.39 | | | | Ref. | 0.55-2.44 | 0.70 | | | |
| ≥30 | 0.82 | | | | | | 1.16 | | | | | |
| ECOG | | | | | | | | | | | | |
| PS 0 | Ref. | 0.71 – 1.36 | 0.91 | 0.93 | 0.67 – 1.28 | 0.64 | Ref. | 0.66-1.53 | 0.99 | 0.93 | 0.60 –1.44 | 0.75 |
| ≥ PS 1 | 0.98 | | | | | | 1.00 | | | | | |
| Disease stage | | | | | | | | | | | | |
| IIIa | Ref. | 0.82 – 1.54 | 0.46 | | | | Ref. | 1.03-2.48 | 0.04 | 1.56 | 0.99 –2.47 | 0.06 |
| IIIb | 1.12 | 0.39 – 1.44 | 0.39 | | | | 1.60 | 0.44-2.07 | 0.90 | 1.01 | 0.43 –2.37 | 0.98 |
| other (IIIc or unknown) | 0.75 | | | | | | 0.95 | | | | | |
| Histology | | | | | | | | | | | | |
| non-squamous | Ref. | 0.66 – 1.26 | 0.58 | | | | Ref. | 0.67-1.54 | 0.94 | | | |
| squamous | 0.91 | | | | | | 1.02 | | | | | |

Abbreviations: cCRT, concurrent chemoradiotherapy; sCRT, sequential chemoradiotherapy; BMI, Body Mass Index; ECOG PS, Eastern Cooperative Group performance status.

*Overall survival*

In the cCRT cohorts, the median OS (mOS) was not reached versus 39.5 months (95%CI 30.1-52.1 months) in patients treated with durvalumab and without durvalumab (Figure 3). The 12-months OS rate for patients treated with durvalumab was higher than for patients treated without durvalumab (83.9% versus 77.5%). The unadjusted HR 0.69 (95%CI, 0.43 – 1.10) and adjusted HR 0.71 (95%CI, 0.44 – 1.13) for OS were better for patients treated with durvalumab than for patients treated without durvalumab, although these results were not significant. None of the other included variables were significantly associated with OS in the multiple Cox regression model (Table 3).

In the sCRT cohorts, the mOS for durvalumab patients was not reached and the mOS for the historical control patients was 24.3 months. The 12-months OS rate for patients treated with durvalumab was higher than for patients treated without durvalumab (93.8% versus 65.8%). The unadjusted HR 0.31 (95%CI, 0.10 – 1.01) and adjusted 0.32 (95%CI, 0.09 – 1.03) for OS were better for patients treated with durvalumab than for patients treated without durvalumab, although these results were not significant. None of the other included variables were significantly associated with OS in the multiple Cox regression model (Table 3).

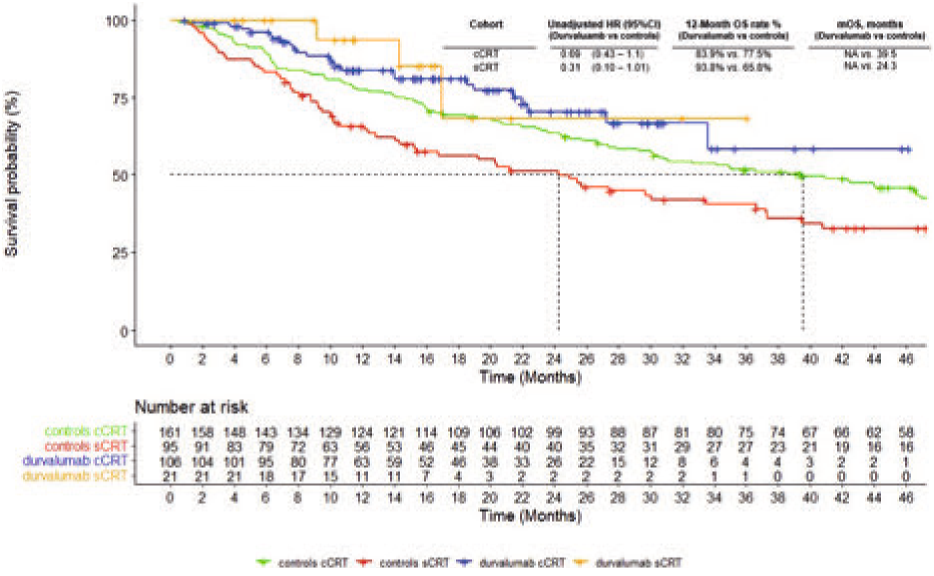


**Figure 3.** Kaplan-Meier curves for overall survival for patients who received cCRT or sCRT with and without durvalumab.

**Table 3.** Overall survival. Simple- and multiple cox-regression model for the cCRT and sCRT population.

| **Overall survival** | **cCRT (n = 267)** | | | | | | **sCRT (n = 116)** | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Simple model** | | | **Multiple model** | | | **Simple model** | | | **Multiple model** | | |
| | *HR* | *95% CI* | *p-value* | *HR* | *95% CI* | *p-value* | *HR* | *95% CI* | *p-value* | *HR* | *95% CI* | *p-value* |
| Type of treatment | | | | | | | | | | | | |
| Control | Ref. | | | | | | Ref. | | | | | |
| Durvalumab | 0.69 | 0.43 - 1.10 | 0.12 | 0.71 | 0.44 – 1.13 | 0.15 | 0.31 | 0.1–1.01 | 0.05 | 0.32 | 0.09 - 1.03 | 0.06 |
| Age | | | | | | | | | | | | |
| <75 years | Ref. | | | | | | Ref. | | | | | |
| ≥ 75 years | 1.29 | 0.77 – 2.16 | 0.33 | 1.19 | 0.67 – 2.11 | 0.55 | 0.87 | 0.50-1.49 | 0.60 | 0.85 | 0.45 - 1.61 | 0.61 |
| Gender | | | | | | | | | | | | |
| Male | Ref. | | | | | | Ref. | | | | | |
| Female | 0.95 | 0.67 – 1.36 | 0.80 | 0.91 | 0.62 – 1.33 | 0.62 | 1.37 | 0.84-2.24 | 0.21 | 1.44 | 0.82 – 2.51 | 0.20 |
| BMI | | | | | | | | | | | | |
| <30 | Ref. | | | | | | Ref. | | | | | |
| ≥30 | 0.74 | 0.43 – 1.28 | 0.28 | | | | 1.15 | 0.54-2.41 | 0.72 | | | |
| ECOG | | | | | | | | | | | | |
| PS 0 | Ref. | | | | | | Ref. | | | | | |
| ≥ PS 1 | 1.47 | 1.01 – 2.13 | 0.05 | 1.43 | 0.97 – 2.10 | 0.07 | 0.96 | 0.58-1.61 | 0.89 | 0.93 | 0.54 - 1.60 | 0.78 |
| Disease stage | | | | | | | | | | | | |
| IIIa | Ref. | | | | | | Ref. | | | | | |
| IIIb | 1.11 | 0.77 – 1.61 | 0.57 | | | | 1.10 | 0.66-1.83 | 0.72 | 1.08 | 0.61 – 1.92 | 0.79 |
| other (IIIc or unkown) | 0.89 | 0.42 – 1.86 | 0.75 | | | | 0.30 | 0.07-1.25 | 0.10 | 0.41 | 0.09 – 1.83 | 0.24 |
| Histology | | | | | | | | | | | | |
| non-squamous | Ref. | | | | | | Ref. | | | | | |
| squamous | 0.86 | 0.48 – 1.57 | 0.62 | | | | 0.95 | 0.72-1.26 | 0.74 | | | |

Abbreviations: cCRT, concurrent chemoradiotherapy; sCRT, sequential chemoradiotherapy; BMI, Body Mass Index; ECOG PS, Eastern Cooperative Group performance status.

## Real-world versus PACIFIC trial

*Durvalumab arm comparison*

A total of 476 patients in the PACIFIC study and 105 patients in our real-world cohort were treated with cCRT followed by durvalumab. In the real-world, the proportion of males was significantly lower (50.9% vs 70.2%) and there were fewer patients with an ECOG PS of 1 (38.7% vs. 50.4%) than in the PACIFIC trial. The proportion of patients with disease stage IIIa (41.5% vs 52.9%) and with squamous histology (30.2% vs. 47.1%) were significantly lower in the real-world than in the PACIFIC trial. Besides, all patients in the PACIFIC trial were treated with durvalumab 42 days after the end of CRT while only 56 patients (52.8%) in the real-world received durvalumab within this timeframe (Table S1 supplementary appendix).

The mPFS observed in the real-world was significantly longer than the mPFS reported in the PACIFIC trial, respectively 27.1 months (95%CI, 19.2 – NR) and 16.8 months (95%CI, 13.0 – 18.1) (HR = 0.64; 95%CI 0.46 – 0.91). The 12-months PFS rate was also higher for patients treated in the real-world versus patients treated in the PACIFIC trial (70.2% versus 55.9%) (Figure 4). For both cohorts, the mOS was not reached and the 12-months survival rates were comparable, 83.8% for the real-world and 83.1% for the trial population. There was no significant difference observed in OS (HR = 0.83; 95%CI 0.54 – 1.26) (Figure 5).

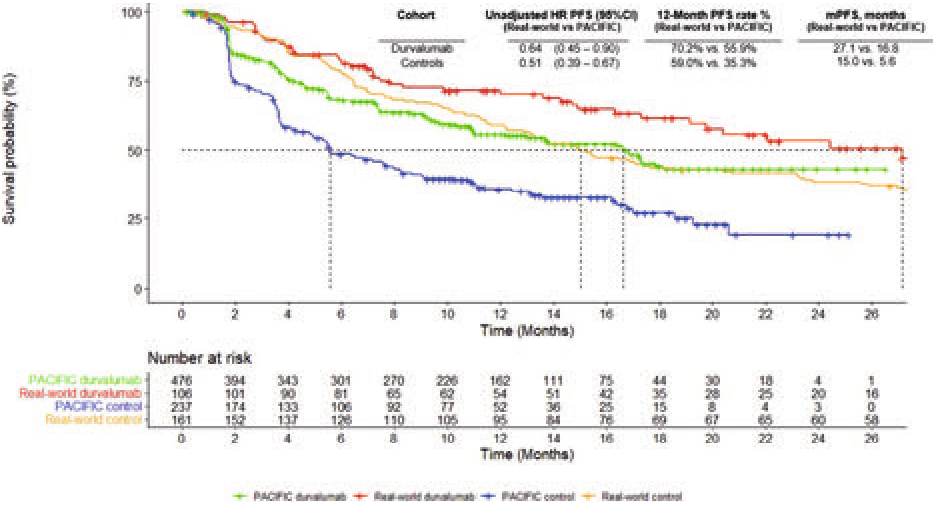


**Figure 4.** Kaplan-Meier curves for progression-free survival of patients who received cCRT with durvalumab and without durvalumab (controls) in the real-world and in the PACIFIC trial.

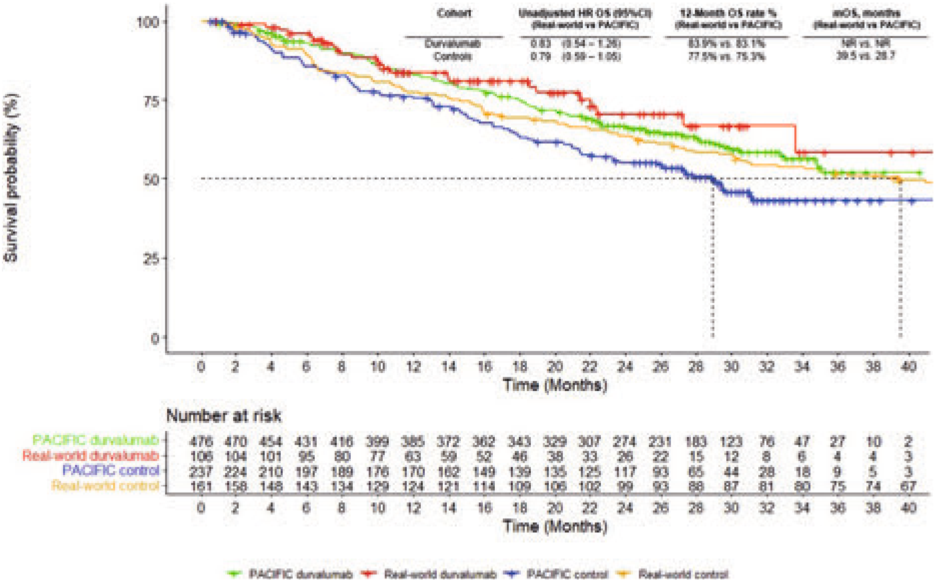


**Figure 5.** Kaplan-Meier curves for overall survival of patients who received cCRT with durvalumab and without durvalumab (controls) in the real-world and in the PACIFIC trial.

*Control arm comparison*

In the PACIFIC study a total of 273 patients were treated with cCRT followed by placebo (hereafter referred to as controls) and in our real-world cohort 161 patients were treated with cCRT alone (historical controls). In the real-world, the proportion of males was lower (53.4% vs 70.0%) and there were fewer patients with an ECOG PS of 1 (30.4% vs. 51.5%) than in the PACIFIC trial. In addition, in the real-world 6 (3.7%) patients with an ECOG PS of 2 were included, while these patients were not included in the PACIFIC trial (Table S1 supplementary appendix).

Furthermore, the mPFS observed in the real-world was significantly longer than the mPFS reported in the PACIFIC trial (15.0 months [95%CI, 12.3 – 20.7] verus 5.6 months [95%CI, 4.6 – 7.8]; HR = 0.51; 95%CI 0.39 – 0.67)). The 12-months PFS rate was also higher for patients treated in the real-world versus patients treated in the PACIFIC trial (59.0 versus 35.3%), see Figure 4. The mOS observed in the real-world was longer than the mOS observed in the PACIFIC (39.5 [95%CI versus 28.7 months [95%CI, 22.9 – NR]) (HR = 0.79; 95%CI 0.59 – 1.05)). The 12-months OS rates were comparable between the real-world and the PACIFIC trial (77.5 versus 75.3%) (Figure 5).

## Discussion

In this retrospective follow-up study, we assessed the real-world effectiveness of durvalumab after cCRT or sCRT in patients with stage III NSCLC. Our findings illustrate that administering durvalumab after CRT improves both PFS and OS compared to historical controls treated with CRT alone, although statistical significance is lacking. Moreover, the mPFS for durvalumab treatment reported in our study (27.1 months) was significantly longer than the mPFS in the PACIFIC trial (16.8 months), while the OS results did not differ. The difference in mPFS could be the result of differences in follow-up methods between the real-world and clinical trial.

For patients treated with cCRT in our cohort, durvalumab improved the PFS (HR 0.74) and the OS (HR 0.71). Our HRs were in line with the HRs reported by Sanker et. al. and Pichert et. al.[25, 26] According to the large study of Sanker et. al (n = 1995), durvalumab after cCRT improved the PFS (HR 0.62, 95% CI 0.55–0.70) and OS (HR 0.57, 95% CI 0.50–0.66) compared to cCRT alone. According to the cancer registry study of Pichert et. al., durvalumab after CRT improved the OS compare to CRT alone (HR 0.71, 95%CI 0.67 – 0.82). However, this result was not stratified for cCRT or sCRT and PFS was not assessed.

For patient treated with sCRT in our study, durvalumab improved both the PFS (HR 0.71) and OS (HR 0.31), although few patients (n=21) were treated with durvalumab. To our knowledge, no studies have evaluated the relative effectiveness of durvalumab after sCRT. Nevertheless, in a large single arm study (PACIFIC-R), patients treated with durvalumab after sCRT had shorter PFS than patients treated with durvalumab after cCRT (19.3 vs 23.7 months). Also the phase II PACIFIC 6 trial reported a short mPFS of 10.9 months [36]. However, in the study of Vranker et.al., the PFS of durvalumab was not influenced by the type of CRT (sequential or concurrent) in the univariable analysis [27]. Of note, sCRT is proposed for elderly and/or less fit patients with clinically relevant comorbidities, which might explain the inferior PFS outcomes of patients treated with sCRT versus patients treated with cCRT [5, 37].

We also compared the PFS of cCRT patients treated with durvalumab in our cohort with the PFS of cCRT patients treated with durvalumab in the trial. We observed a significant HR of 0.64 (95%CI 0.46 – 0.91) in favor of durvalumab in the real-world. This additional PFS benefit in the real-world population does not extend to OS outcomes because the OS benefit in our cohort was comparable to the OS reported in the PACIFIC trial. Our observation is in line with the results of the PACIFIC-R study, which showed a longer PFS (21.7 months) [10] than in the PACIFIC trial (16.8 months) [3]. Bruni et al. also found superior PFS outcomes (23.0 months) in real-world patients receiving durvalumab [22]. The OS observed in both studies did not reach the median and therefore could not be compared with the registration trial.

These findings give rise to the question whether the PFS observed in the real world is overestimated, especially in the context of a comparable OS. This observation may be explained by differences in establishing progressive disease. First, in the real-world, the frequency of radiological imaging is lower and less consistent than the strictly timed imaging scheme (every 8 weeks) used in the PACIFIC trial, which caused delayed or even missed progression events near the end of follow up. [38, 39] Second, the radiological evaluation of treatment response in the context of CRT is complex because it is difficult to distinguish true progression from radiation fibrosis [40]. In clinical practice, physicians may be extra careful to declare that patients have progressive disease and to prematurely discontinue immunotherapy. Third, contrary to the original RECIST criteria, the iRECIST criteria require confirmation of progressive disease within 4-8 weeks after the initial signs of progression of new disease [41-44] . So, the use of the iRECIST in the real world could delay the detection of progression events and probably fewer true progression events may occur. [42] These effects might be observed in the decline around two months in PFS KM curves of the PACIFC trial (Figure 4), whereas the first drop in the PFS curve of our cohort is later and smaller than the first drop in the PFS curve of the PACIFIC. Our observation of prolonged PFS but similar OS in the real-world versus the trial appears unique to durvalumab post-CRT treatment, since we did not find comparable results for other immunotherapy treatments in the context of NSCLC.

The real world effectiveness of durvalumab appears to be a complex area of research for several reasons. In addition to the aforementioned challenge of defining progressive disease in the setting of CRT and immunotherapy, several studies [11,16,22,29] used the end of CRT rather than the start of durvalumab as the index date for the calculation of PFS which leads to an overestimated PFS. Also, the duration of sCRT is longer than cCRT, which may have affected the outcome measures of durvalumab in studies that did not include historical controls. Finally, including a historical cohort requires the imputation of an index date and exclusion of patients that progressed before that date. The strengths of our study are that we used a valid index date and minimized the risk for immortal time bias. [45] We also included historical controls to investigate the relative effectiveness of durvalumab and we are the first to include a sCRT cohort. Our study also has some limitations. First, our study lacks power to properly evaluate the effect of durvalumab in a cohort of sCRT pretreated patients. Second, the follow-up time for patients treated with durvalumab was relatively short which explains why the median OS was not reached. Another limitation is that we do not have information on the follow-up methods for disease progression to confirm our explanation for the differences seen in PFS between the real-world and the PACIFIC trial. Lastly, at baseline, the tumor PDL1 expression was not available in the majority of patients (n=244, 64%).

## Conclusion

In both the sCRT and cCRT cohort, we observed that durvalumab following CRT improves the survival outcomes compared to CRT alone. The observed PFS of durvalumab after cCRT in our cohort is significantly longer than the PFS reported in the PACIFIC trial which could be the result of differences in follow-up methods between the real-world and clinical trial.

## References

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA. Cancer J. Clin. 71(3), 209–249 (2021).
2. Huber RM, De Ruysscher D, Hoffmann H, Reu S, Tufman A. Interdisciplinary multimodality management of stage III nonsmall cell lung cancer. Eur. Respir. Rev. 28(152) (2019).
3. Antonia SJ, Villegas A, Daniel D, et al. Overall Survival with Durvalumab after Chemoradiotherapy in Stage III NSCLC. N. Engl. J. Med. 377(20), 1919–1929 (2017).
4. Spigel DR, Faivre-Finn C, Gray JE, et al. Five-Year Survival Outcomes From the PACIFIC Trial: Durvalumab After Chemoradiotherapy in Stage III Non–Small-Cell Lung Cancer. J. Clin. Oncol. (2022).
5. Postmus PE, Kerr KM, Oudkerk M, et al. Early and locally advanced non-small-cell lung cancer (NSCLC): ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann. Oncol. 28 (Supplement 4), iv1–iv21 (2017).
6. Ettinger DS, Wood DE, Aisner DL, et al. Non-small cell lung cancer, Version 2.2021 featured updates to the NCCN guidelines. JNCCN J. Natl. Compr. Cancer Netw. 19(3), 254–266 (2021).
7. Cramer-Van Der Welle CM, Peters BJM, Schramel FMNH, Klungel OH, Groen HJM, Van De Garde EMW. Systematic evaluation of the efficacy-effectiveness gap of systemic treatments in metastatic nonsmall cell lung cancer. Eur. Respir. J. 52(6) (2018).
8. Cramer-van der Welle CM, Verschueren MV., Tonn M, et al. Real-world outcomes versus clinical trial results of immunotherapy in stage IV non-small cell lung cancer (NSCLC) in the Netherlands. Sci. Rep. 11(1), 1–9 (2021).
9. Aupérin A, Le Péchoux C, Rolland E, et al. Meta-analysis of concomitant versus sequential radiochemotherapy in locally advanced non - small-cell lung cancer. J. Clin. Oncol. 28(13), 2181–2190 (2010).
10. Girard N, Smit H, Sibille A, et. al. Treatment characteristics and real-world progression-free survival in patients with unresectable stage III NSCLC who received durvalumab after chemoradiotherapy: Findings from the PACIFIC-R study. J. Thorac. Oncol. (2022). Available at: https://www.sciencedirect.com/science/article/pii/S1556086422018536
11. Taugner J, Käsmann L, Eze C, et al. Durvalumab after Chemoradiotherapy for PD-L1 Expressing Inoperable Stage III NSCLC Leads to Significant Improvement of Local-Regional Control and Overall Survival in the Real-World Setting. Cancers (Basel). 13(7), 1613 (2021).
12. Shaverdian N, Thor M, Shepherd AF, et al. Radiation pneumonitis in lung cancer patients treated with chemoradiation plus durvalumab. Cancer Med. 9(13), 4622–4631 (2020).
13. Jegannathen A. Real-world data of using durvalumab in stage III non-small cell lung cancer (NSCLC): West Midlands experience. Lung Cancer. 139(2020), S51 (2020).
14. Jain P, Murray P, Clarke K, et al. Early experience of maintenance durvalumab post chemoradiation (CRT) in stage III non-small cell lung cancer (NSCLC) across West Yorkshire network: from Expanded Access Programme (EAP) to routine clinical use. Lung Cancer. 139(2020), S46 (2020).
15. Faehling M, Schumann C, Christopoulos P, et al. Durvalumab after definitive chemoradiotherapy in locally advanced unresectable non-small cell lung cancer (NSCLC): Real-world data on survival and safety from the German expanded-access program (EAP). Lung Cancer. 150, 114–122 (2020).
16. Desilets A, Blanc-Durand F, Lau S, et al. Durvalumab therapy following chemoradiation compared with a historical cohort treated with chemoradiation alone in patients with stage III non–small cell lung cancer: A real-world multicentre study. Eur. J. Cancer. 142, 83–91 (2021).
17. Chu C, Chiu T, Wang C, et.al. Consolidation treatment of durvalumab after chemoradiation in real-world patients with stage. Thorac. Cancer. 11(6), 1541–1549 (2020).

18. Tsukita Y, Yamamoto T, Mayahara H, et al. Intensity-modulated radiation therapy with concurrent chemotherapy followed by durvalumab for stage III non-small cell lung cancer: A multi-center retrospective study. Radiother. Oncol. 160(2021), 266–272 (2021).
19. LeClair JN, Merl MY, Cohenuram M, Luon D. Real-World Incidence of Pneumonitis in Patients Receiving Durvalumab. Clin. Lung Cancer. 23(1), 34–42 (2022).
20. Miura Y, Mouri A, Kaira K, et.al. Chemoradiotherapy followed by durvalumab in patients with unresectable advanced non-small. Thorac. Cancer. 11(5). 1280–1287 (2020).
21. Offin M, Shaverdian N, Rimner A, et al. Clinical outcomes, local–regional control and the role for metastasis-directed therapies in stage III non-small cell lung cancers treated with chemoradiation and durvalumab. Radiother. Oncol. 149, 205–211 (2020).
22. Bruni A, Scotti V, Borghetti P, et al. A Real-World, Multicenter, Observational Retrospective Study of Durvalumab After Concomitant or Sequential Chemoradiation for Unresectable Stage III Non-Small Cell Lung Cancer. Front. Oncol. 11, 3854 (2021).
23. Huang Y, Zhao JJ, Soon YY, et al. Real-world experience of consolidation durvalumab after concurrent chemoradiotherapy in stage III non-small cell lung cancer. Thorac. cancer. (2022). 13(11), 3152-3161 (2022).
24. Ohri N, Halmos B, Bodner WR, et al. Who Benefits the Most From Adjuvant Durvalumab After Chemoradiotherapy for Non-small Cell Lung Cancer? An Exploratory Analysis. Pract. Radiat. Oncol. 11(2), e172–e179 (2021).
25. Sankar K, Bryant AK, Strohbehn GW, et al. Real World Outcomes versus Clinical Trial Results of Durvalumab Maintenance in Veterans with Stage III Non-Small Cell Lung Cancer. Cancers (Basel). 14(3) (2022).
26. Pichert MD, Canavan ME, Maduka RC, et al. Immunotherapy After Chemotherapy and Radiation for Clinical Stage III Lung Cancer. JAMA Netw. Open. 5(8), 1–13 (2022).
27. Vrankar M, Stanic K, Jelercic S, Ciric E, Vodusek AL, But-Hadzic J. Clinical outcomes in stage III non-small cell lung cancer patients treated with durvalumab after sequential or concurrent platinum-based chemoradiotherapy - Single institute experience. Radiol. Oncol. 55(4), 482–490 (2021)
28. Wang CC, Chiu LC, Ju JS, et al. Durvalumab as consolidation therapy in post-concurrent chemoradiation (Ccrt) in unresectable stage iii non-small cell lung cancer patients: A multicenter observational study. Vaccines. 9(10) (2021).
29. Jung HA, Noh JM, Sun JM, et al. Real world data of durvalumab consolidation after chemoradiotherapy in stage III non-small-cell lung cancer. Lung Cancer. 146, 23–29 (2020).
30. Park C-K, Jeon N, Park H-K, et al. A propensity-matched retrospective comparative study with historical control to determine the real-world effectiveness of durvalumab after concurrent chemoradiotherapy in unresectable stage III non-small cell lung cancer. Cancers. 15(5), 1606 (2023).
31. Wang Y, Zhang T, Huang Y, et al. Real-world Safety and Efficacy of Consolidation Durvalumab after Chemoradiotherapy for Stage III Non-small-cell Lung Cancer: A Systematic Review and Meta-analysis. Int. J. Radiat. Oncol. Biol. Phys. 112(5), 1154-1164 (2022).
32. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)-A metadata-driven methodology and workflow process for providing translational research informatics support. J. Biomed. Inform. 42(2), 377–381 (2009).
33. Suissa S. Immortal time bias in pharmacoepidmiology. Am. J. Epidemiol. 167(4), 492-499 (2008).
34. Antonia SJ, Villegas A, Daniel D, et al. Overall Survival with Durvalumab after Chemoradiotherapy in Stage III NSCLC. N. Engl. J. Med. 379(24), 2342–2350 (2018).
35. Guyot P, Ades AE, Ouwens MJNM, Welton NJ. Enhanced secondary analysis of survival data: Reconstructing the data from published Kaplan-Meier survival curves. BMC Med. Res. Methodol. 12 (2012).

36. Garassino MC, Mazieres J, Reck M, et al. Durvalumab after sequential chemoradiotherapy in stage III, unresectable NSCLC: The phase 2 PACIFIC-6 trial. J. Thorac. Oncol. 17(12). 1415-1427 (2022).
37. Belderbos J, Uitterhoeve L, Van Zandwijk N, et al. Randomised trial of sequential versus concurrent chemoradiotherapy in patients with inoperable non-small cell lung cancer. Eu. J. Cancer. 43(1), 114-121 (2007).
38. Panageas KS, Ben-Porat L, Dickler MN, Chapman PB and Schrag D. When you look matters: The effect of assessment schedule on progression-free survival. J. Natl. Cancer Inst. 99(6), 428–432 (2007).
39. Adamson BJS, Ma X, Griffith SD, Sweeney EM, Sarkar S and Bourla AB. Differential frequency in imaging-based outcome measurement: Bias in real-world oncology comparative-effectiveness studies. Pharmacoepidemiol. Drug Saf. 31(1), 46–54 (2022)
40. Shukla NA, Hanna NH. Practical challenges in patients with stage III NSCLC receiving checkpoint inhibitors after chemoradiation. Lung Cancer Manag. 9(1), 10–13 (2020).
41. Luis Ramon-Patino J, Schmid S, Lau S, et al. Open access iRECIST and atypical patterns of response to immuno-oncology drugs. J. Immunother. Cancer. 10, 4849 (2022)
42. Park HJ, Kim GH, Kim KW, et al. Comparison of RECIST 1.1 and irecist in patients treated with immune checkpoint inhibitors: A systematic review and meta-analysis. Cancers (Basel). 13(1), 1–14 (2021).
43. Seymour L, Bogaerts J, Perrone A, et al. iRECIST: guidelines for response criteria for use in trials testing immunotherapeutics. Lancet Oncol. 18(3), 143-152 (2017)
44. Persigehl T, Lennartz S and Schwartz L. iRECIST: how to do it. Cancer imaging. 20(1). (2020)
45. Strom BL, Kimmel SE and Hennessy S. Textbook of pharmacoepidemiology. 3rd ed. John Wiley & Sons, Hoboken.

## Supplementary

**Table S1.** Patient characteristics of NSCLC patients who received cCRT with or without durvalumab in the real-world and in the PACIFIC trial.

| **Characteristics** | | **Durvalumab** | | | **Controls** | | |
|---|---|---|---|---|---|---|---|
| | | **Real-world** | **PACIFIC** | **p-value** | **Real-world** | **PACIFIC** | **P-value** |
| | | **(n = 106)** | **(n = 476)** | | **(n=161)** | **(n = 237)** | |
| Age (median(range)) | | 65 (36 - 83) | 64 (31-84) | - | 65 (37 - 83) | 64 (23-90) | - |
| Sex (male, n (%)) | | 54 (50.9) | 334 (70.2) | <0.01 | 86 (53.4) | 166 (70.0) | 0.19 |
| BMI (mean, sd) | | 26.0 (4.5) | - | - | 24.8 (4.5) | - | - |
| ECOG PS (n (%)) | | | | 0.03 | | | <0.01 |
| | 0 | 62 (58.5) | 234 (49.2) | | 90 (55.9) | 114 (48.1) | |
| | 1 | 41 (38.7) | 240 (50.4) | | 49 (30.4) | 122 (51.5) | |
| | ≤2 | 1 (0.9) | - | | 6 (3.7) | - | |
| | unknown | 2 (1.9) | - | | 16 (9.9) | - | |
| Disease stage (n, (%)) | | | | <0.01 | | | 0.43 |
| | IIIa | 44 (41.5) | 252 (52.9) | | 83 (51.6) | 125 (52.7) | |
| | IIIb | 46 (43.4) | 212 (44.5) | | 70 (43.5) | 107 (45.1) | |
| | IIIc | 6 (5.7) | 0 | | 3 (1.9) | 0 | |
| | nos | 10 (9.4) | 12 (2.5) | | 5 (3.1) | 5 (2.1) | |
| Histology (n, (%)) | | | | <0.01 | | | 0.16 |
| | Squamous | 32 (30.2) | 224 (47.1) | | 58 (36.0) | 102 (43.0) | |
| | Nonsquamous | 74 (69.8) | 252 (52.9) | | 103 (64.0) | 135 (57.0) | |
| Days between end radiation and start durvalumab | | | | - | | | - |
| | ≤ 42 days | 56 (52.8) | 476 (100%) | | - | - | |
| | > 42 days | 50 (47.2) | 0 | | - | - | |

Abbreviations: ECOG PS, Eastern Cooperative Group performance status; PD-L1, programmed death ligand-1; nos, not otherwise specified

# 2.2

# Pembrolizumab plus chemotherapy per PD-L1 stratum in patients with metastatic non-small cell lung cancer: real-world effectiveness versus trial efficacy.

Marjon V Verschueren, Bas JM Peters, Lourens T Bloem, Veerle R Kruik,
Elien B Uitvlugt, Anette R Bijsmans, Antoine CG Egberts,
Ewoudt MW van de Garde

*Clinical Lung Cancer. 2024 Mar;25(2):119-127*
*DOI:10.1016/j.cllc.2023.12.011.*

# Abstract

## Background

Clinical trial efficacy and real-world effectiveness of oncological treatments can differ. This study assessed the real-world survival outcomes of first-line pembrolizumab plus chemotherapy per PD-L1 stratum in patients with metastatic non-small cell lung cancer (mNSCLC) and compared them to clinical trial results.

## Methods

All patients with non-squamous and squamous mNSCLC who received first-line pembrolizumab plus chemotherapy in seven Dutch teaching hospitals between 01-01-2019 and 31-12-2021 were included. Hazard ratios (HR) with confidence intervals (95%CI) for overall survival (OS) and progression-free survival (PFS) were estimated to determine the efficacy-effectiveness gap (EE gap) between real-world and clinical trial, stratified by PD-L1 stratum.

## Results

The non-squamous cohort (n=486) consisted of 269 patients with PD-L1 <1%, 158 with PD-L1 1-49%, and 59 with PD-L1 ≥50%. The squamous cohort (n=117) consisted of 70 patients with PD-L1 <1% and 47 with PD-L1 ≥1%. For OS, an EE gap was observed in non-squamous patients with PD-L1 <1% (HR 1.38 (95%CI 1.06 – 1.78; median OS (mOS) 10.0 versus 17.2 months) and HRs consistently >1 in all other non-squamous and squamous PD-L1 strata, although not statistically significant. No EE-gap for PFS was observed in any stratum.

## Conclusion

No significant efficacy-effectiveness gap was found for pembrolizumab plus chemotherapy, except in the stratum non-squamous metastatic NSCLC with <1% PD-L1 tumor expression. In these patients, the survival in real-world was considerably shorter compared to the clinical trial results. Further studies are needed to determine which patient, treatment and or context factors contribute to this disparity.

## Introduction

Approximately one million patients are worldwide diagnosed with metastatic non-small-cell lung cancer (NSCLC) each year [1]. The introduction of immune checkpoint inhibitors targeting PD-1/PD-L1, such as pembrolizumab, nivolumab, atezolizumab and cemiplimab, has improved the survival outcomes for metastatic NSCLC patients without oncogenic driver mutations [2].

The KEYNOTE-189 and KEYNOTE-407 trials evaluated the use of pembrolizumab in combination with chemotherapy as a first-line treatment option for metastatic NSCLC patients [3,4]. Both trials demonstrated that pembrolizumab plus chemotherapy improved survival outcomes compared to chemotherapy alone, regardless of PD-L1 expression. In addition, for patients with a high PD-L1 expression (PD-L1 ≥50%) the KEYNOTE-042 trial reported improved survival outcomes with pembrolizumab as monotherapy [5]. As a result, current guidelines [6] recommend pembrolizumab plus chemotherapy in patients with no (PD-L1 <1%) or intermediate (PD-L1 1-49%) PD-L1 expression. For patient with high PD-L1 expression (PD-L1 ≥ 50%), pembrolizumab monotherapy is recommended, unless a rapid tumor response is required, in which pembrolizumab plus chemotherapy may be considered.

Treatment effects observed in clinical trials may differ from effects observed in clinical practice, also known as the efficacy-effectiveness gap (EE gap) [7,8]. Recently, a systematic evaluation of the EE-gap of systemic treatments in metastatic NSCLC (n=1063) revealed a median OS (mOS) being approximately 30% shorter in clinical practice compared to the corresponding clinical trials [8]. This EE gap was consistently observed for all chemotherapy regimens studied. Now that pembrolizumab is added to chemotherapy, a similar EE gap may exist. Because the choice of treatment regimen is based on PD-L1 status, an appropriate assessment of an EE-gap should be conducted per PD-L1 stratum to prevent selection bias.

The objective of the present study is to assess the real-world survival outcomes of pembrolizumab plus chemotherapy per PD-L1 stratum of patients with non-squamous and squamous metastatic NSCLC and compare these outcomes to the clinical trial results.

## Methods

### Study design, setting, participants

This study was a multicenter cohort study conducted in the Santeon group, a network of seven large non-university teaching hospitals spread geographically over the Netherlands and serving approximately 15% of the Dutch population. Eligible patients were identified in the ‘Santeon Farmadatabase’, a database that contains prospectively documented individual patient data on drug prescriptions and diagnoses for all patients treated within one of the participating hospitals [9]. For this study, patients were selected if they were diagnosed with metastatic NSCLC without EGFR or ALK genomic aberrations and had received at least one cycle of pembrolizumab plus chemotherapy as first line treatment between 1 January 2019 and 31 December 2021. The index date was defined as the date of the first pembrolizumab administration. Subsequently, the electronic health records from all selected patients were manually reviewed to collect baseline characteristics (30 days before and after diagnosis), treatment characteristics and follow-up outcomes. The last date of follow-up was 1 August 2022. The PD-L1 status, measured through immunohistochemistry, was collected from the EHR, and patients without an assessable PD-L1 status were excluded at this stage. To compare findings with the relevant corresponding KEYNOTE clinical trials, the study population was split into two cohorts. The non-squamous cohort included patients with non-squamous histological subtypes who received pembrolizumab plus pemetrexed-platinum. The squamous cohort included patients with squamous or unknown histological subtypes who received pembrolizumab plus paclitaxel-platinum. The construction of the cohorts is schematically depicted in Figure 1. Patients were stratified based on PD-L1 status within both cohorts, allowing for a comparison with the corresponding strata in the trials.

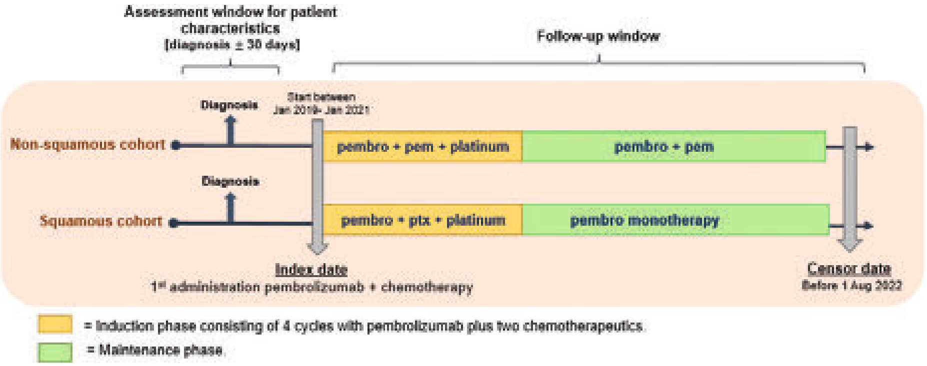


**Figure 1.** Study design

Abbreviations: ptx, paclitaxel; pem, pemetrexed; pembro, pembrolizumab; platinum, cisplatin or carboplatin.

## Outcomes

The primary outcome measures were overall survival (OS) and progression-free survival (PFS). The OS was calculated from the index date till the date of death or censoring. Date of censoring was the last clinic visit before the end of follow-up (1 August 2022). PFS was calculated from the index date till the date of disease progression, death or censoring, whichever occurred first. Progression of disease was determined according to local protocols including semi-standardized response assessment schemes, where the first radiological response assessment was scheduled 6-9 weeks after treatment initiation and whenever feasible (immune) RECIST criteria were applied. In this study, the date of disease progression was defined as the date of the first medical note of the thoracic oncologist stating that the disease had progressed.

### Patient and treatment characteristics

The following patient and treatment characteristics were collected for every patient; age at diagnosis, sex, body mass index (BMI), Eastern Cooperative Oncology Group-Performance status (ECOG PS), histological subtype, presence of brain metastasis, the type of platinum-based chemotherapy (cisplatin or carboplatin), the number of cycles of platinum based chemotherapy, the treatment duration of pembrolizumab, and whether or not further lines of treatment have been initiated.

### Data analysis

Statistical software (R version 4.1.2) was used to conduct the data analysis. The patient and treatment characteristics were presented and compared between the real-world PD-L1 strata within the non-squamous and squamous cohorts, using chi-square, t-tests, Wilcoxson sum rank test and fisher exact tests when appropriate. Notably, in absence of PD-L1 stratified patient data from the clinical trials, a comparison with characteristics of trial participants was only possible with the overall trial population.

The Kaplan-Meier (KM) method was used to visualize the PFS and OS survival curves and estimate median OS (mOS) and median PFS (mPFS) of real-world patients for the following strata: non-squamous <1% PD-L1, non-squamous 1-49% PD-L1, non-squamous ≥50% PD-L1, squamous <1% PD-L1, and squamous >1% PD-L1.

The EE-gap for the survival curves for the aforementioned strata of the real-world population were compared with the data from the corresponding trial strata using Cox proportional regression modelling to estimate Hazard Ratios and 95% confidence intervals (95%CI). The algorithm developed by Guyot and colleagues [10] was used to reconstruct data from the available OS and PFS curves from the PD-L1 strata of the KEYNOTE-189 [11] and -407 [12] trials. The survival curves of all models were visually assessed to test whether the proportional hazard assumption was met, and in cases where the assumption was violated, HRs were not reported. Lastly, a sensitivity analysis was performed, including only patients with an ECOG PS of 0 and 1, to evaluate whether differences in ECOG PS impacted the OS and PFS EE gaps.

### Ethical statement

All methods were carried out in accordance with relevant guidelines and regulations. The ethics committee—the Santeon Institutional Review Board—approved the study (SDB 2019-008) and waived the need for informed consent because of the retrospective nature of the study and most patients were deceased at time of conducting the study. The study was performed in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

## Results

In total, 633 patients started with pembrolizumab plus chemotherapy in the study period. Of these patients, 30 (23 non-squamous and 7 squamous) had no PD-L1 assessment available, resulting in an overall study population of 603 patients, including 486 patients with non-squamous and 117 patients with squamous histology.

### Patient and treatment characteristics

Table 1 presents the patient characteristics and Table 2 presents the treatment characteristics of both cohorts and their PD-L1 strata.

*Non-squamous cohort*

In the non-squamous population, 269 (55.4%) patients had <1% PD-L1 expression, 158 (32.5%) patients had 1-49% PD-L1 expression, and 59 patients (12.1%) had ≥50% PD-L1 expression. In the KEYNOTE-189 trial 410 patients received pembrolizumab plus chemotherapy of whom 127 (31.0%) patients had <1% PD-L1 expression, 128 (31.2%) patients had 1-49% PD-L1 expression and 132 (32.2%) patients had ≥50% PD-L1 expression. The characteristics of the patients were similar for all PD-L1 strata in clinical practice, except for the presence of brain metastases and the duration of pembrolizumab therapy.

In comparison to the KEYNOTE-189, the proportion of patients with an ECOG PS ≥ 2 and an unknown ECOG PS were higher in all three PD-L1 strata (Table 1). The proportion of patients receiving ≥ 4 cycles of platinum-based chemotherapy and subsequent treatment were lower in all PD-L1 strata (Table 2).

*Squamous cohort*

In the squamous cohort, 70 (59.8%) and 47 (40.2%) patients had a PD-L1 expression of <1% and ≥1% respectively, compared to 95 (34.2%) and 176 (63.4%) patients in pembrolizumab arm of the KEYNOTE-407 trial (n =278). In the real-world, brain metastasis were more frequent in patients with ≥1% PD-L1 expression than in patients with <1% PD-L1 expression.

In comparison to the KEYNOTE-407, the median age and the proportion of patients with ECOG PS ≥ 2 and an unknown ECOG PS were higher in both PD-L1 strata. The proportions of patients receiving ≥ 4 cycles of platinum-based chemotherapy and subsequent treatment were lower in both PD-L1 strata. (Table 2).

**Table 1.** Patient characteristics stratified per PD-L1 status and overall characteristics from the KEYNOTE trials for non-squamous and squamous metastatic NSCLC patients.

| Characteristics | Non-squamous | | | | | Squamous | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Real-world (n = 486) | | | | KEYNOTE-189 | Real-world (n =117) | | | KEYNOTE-407 |
| | PD-L1 < 1% | PD-L1 1-49% | PD-L1 ≥50% | p-value* | Pembrolizumab arm | PD-L1 <1% | PD-L1 ≥ 1% | p-value* | Pembrolizumab arm |
| | (n = 269) | (n = 158) | (n = 59) | | (n = 410)† | (n = 70) | (n = 47) | | (n = 278)†† |
| Age (median, [range]) | 66 [35-85] | 65 [40-81] | 61 [38-80] | 0.49 | 65 [34 – 84] | 71 [56 – 86] | 68 [53 – 82] | 0.10 | 65 [29 – 87] |
| Sex (male, n (%)) | 139 (51.6) | 66 (41.8) | 27 (45.8) | 0.13 | 254 (62.0) | 51 (72.8) | 32 (68.1) | 0.72 | 220 (79.1) |
| BMI (mean, (sd)) | 24.5 (4.4) | 25.4 (4.8) | 25.5 (4.7) | 0.41 | - | 26.2 (6.0) | 24.9 (3.9) | 0.18 | - |
| ECOG-PS (n (%)) | | | | 0.95 | | | | 0.73 | |
| 0 | 91 (33.8) | 58 (36.7) | 21 (36.7) | | 186 (45.4) | 18 (25.7) | 16 (34.0) | | 73 (26.3) |
| 1 | 93 (34.6) | 55 (34.8) | 18 (30.5) | | 221 (53.9) | 30 (42.9) | 17 (36.2) | | 205 (73.7) |
| ≥2 | 27 (10.0) | 12 (7.6) | 7 (11.8) | | 1 (0.2) | 8 (11.4) | 4 (8.5) | | 0 |
| unknown | 58 (18.6) | 33 (20.9) | 13 (22.0) | | 2 (0.5) | 17 (24.3) | 10 (21.3) | | 0 |
| Histology (n, (%)) | | | | 0.91 | | | | 0.06 | |
| Adeno | 244 (90.7) | 145 (91.8) | 53 (89.8) | | 394 (96.1) | - | - | | - |
| Squamous | - | - | - | | - | 57 (81.4) | 45 (95.7) | | 272 (97.8) |
| NOS or other | 25 (9.3) | 13 (8.2) | 6 (10.2) | | 16 (3.9) | 13 (18.6) | 2 (4.3) | | 6 (2.2) |
| Brain metastases (n, (%)) | 50 (18.6) | 23 (38.9) | 16 (27.1) | <0.01 | 73 (17.8) | 4 (5.7) | 5 (10.6) | <0.01 | 20 (7.2) |

Abbreviations: BMI, Body Mass Index; ECOG-PS, Eastern Cooperative Group performance status; NA, not applicable; nos, not otherwise specified.

† Intention to treat population including all patients who were randomly allocated to pembrolizumab plus chemotherapy

††As treated population which included all patients who underwent randomization and received ≥ 1 dose of pembrolizumab plus chemotherapy

* p-value represents a comparison of real-world patient characteristics between the PD-L1 stratum within the non-squamous or squamous cohort

**Table 2.** Treatment characteristics stratified per PD-L1 status and overall characteristics from the KEYNOTE trials for non-squamous and squamous metastatic NSCLC patients.

| Treatment characteristics | Non-squamous | | | | | Squamous | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Real-world (n = 486) | | | | KEYNOTE-189 | Real-world (n = 117) | | | KEYNOTE-407 |
| | PD-L1 <1% (n = 269) | PD-L1 1-49% (n = 158) | PD-L1 <50% (n = 59) | P value* | Pembrolizumab arm (n =405$^{†}$ /410$^{††}$) | PD-L1 <1% (n =70) | PD- L1 >1% (n = 47) | P value* | Pembrolizumab arm (n = 278) † |
| Type of platinum-based drug (n, (%)) | | | | 0.39 | | | | 0.77 | |
| Carboplatin | 219 (81.4) | 120 (75.9) | 46 (78.0) | | 294 (71.7)$^{†}$ | 60 (85.7) | 42 (89.4) | | 167 (60.1) |
| Cisplatin | 50 (18.6) | 38 (24.1) | 13 (22.0) | | 111 (27.4)$^{†}$ | 10 (14.2) | 5 (10.6) | | 113 (40.6) |
| Number of platinum cycles in the induction phase (n, (%)) | | | | 0.37 | | | | 0.35 | |
| < 4 | 99 (36.7) | 55 (34.8) | 16 (27.1) | | 71 (17.5)$^{†}$ | 28 (40.0) | 14 (29.8) | | 59 (21.2) |
| ≥ 4 | 170 (63.2) | 103 (65.2) | 43 (72.9) | | 334 (82.5)$^{†}$ | 42 (60.0) | 33 (70.2) | | 219 (78.8) |
| Treatment duration pembrolizumab (mean, (sd)) | 6.3 (7.1) | 7.9 (7.6) | 9.9 (10.2) | 0.02 | 9.8 (7.8)$^{†}$ | 6.6 (7.0) | 8.6 (8.1) | 0.02 | 6.3 (4.1) |
| Subsequent therapy | | | | 0.45 | | | | | |
| Any subsequent therapy (n, (%)) | 45 (16.7) | 25 (15.8) | 6 (10.2) | | 183 (44.6)$^{††}$ | 15 (21.4) | 11 (23.4) | 0.98 | 89 (32.0) |
| Number of subsequent therapies (n,% of any subsequent therapy) | | | | 0.08 | | | | <0.01 | |
| 1 | 38 (84.4) | 19 (76.0) | 5 (83.3) | | 107 (58.4)$^{††}$ | 11 (73.3) | 7 (63.6) | | NA |
| 2 | 3 (6.7) | 6 (14.0) | 0 | | 44 (24.0)$^{††}$ | 3 (20.0) | 4 (36.4) | | NA |
| ≥3 | 4 (8.9) | 0 | 1 (16.7) | | 32 (17.5)$^{††}$ | 1 (6.7) | 0 | | NA |

Abbreviations: NA, not applicable; sd, standard deviation.
† Intention to treat population including all patients who were randomly allocated to pembrolizumab plus chemotherapy (non-squamous, n=405; squamous, n =278))
††As treated population which included all patients who underwent randomization and received ≥ 1 dose of pembrolizumab plus chemotherapy (n=410)
* P-value represents a comparison between real-world patient characteristics between the PD-L1 strata within the non-squamous or squamous cohort

## Survival outcomes

*Non-squamous cohort*

For all PD-L1 strata, the mOS in the real-world was shorter than in the KEYNOTE-189 trial (Figure 2). The difference in mOS was most pronounced in patients with a PD-L1 expression of less than 1% with a significantly shorter mOS in the real-world than in the trial (10.0 versus 17.1 months; HR = 1.38 [95%CI, 1.06 – 1.78]). The mOS was slightly, but not significantly, shorter in the real-world for patients with a PD-L1 expression of 1-49% (19.9 versus 21.8 months; HR 1.10 [95%CI, 0.80 – 1.51]) and for patients with a PD-L1 expression of ≥50% (26.2 versus 27.7 months, HR = 1.21 [95%CI, 0.76 – 1.89]). The one-year survival was lower across all PD-L1 strata when compared to the trial, namely 44.8% versus 63.4% for PD-L1 <1%, 59.1% versus 71.1% for PD-L1 1-49% and 63.9% versus 73.3% for PD-L1 ≥ 50%. Regarding PFS, for the PD-L1 <1% and 1-49% strata no significant difference was observed between the real-world and the trial. In the PD-L1 ≥50% stratum, the mPFS was longer in the real-world than in the trial (22.2 versus 11.1 months). However, this should be interpreted with caution since the proportional hazard assumption was violated (Figure 2F). A sensitivity analysis including only patients with ECOG 0 and 1 yielded similar OS and PFS EE-gaps (supplementary Table S1).

*Squamous cohort*

The mOS observed in patients with <1% PD-L1 was shorter in the real-world than in the KEYNOTE-407 trial (10.0 versus 15.0 months; HR = 1.37 [95%CI 0.92 – 2.04]), see Figure 3. For patients with ≥ 1% PD-L1, the mOS observed in the real-world was longer than the mOS observed in the trial (25.6 versus 18.9 months). However, the observed mOS should be interpreted with caution since the proportional hazard assumption was violated (Figure 3B). The one-year survival for PD-L1 <1% was lower in the real-world than in the trial (49.9% versus 64.2%), while it was similar for PD-L1 ≥1% (61.3% versus 64.8%). There were no differences in PFS with the KEYNOTE-407 trial in both PD-L1 strata.

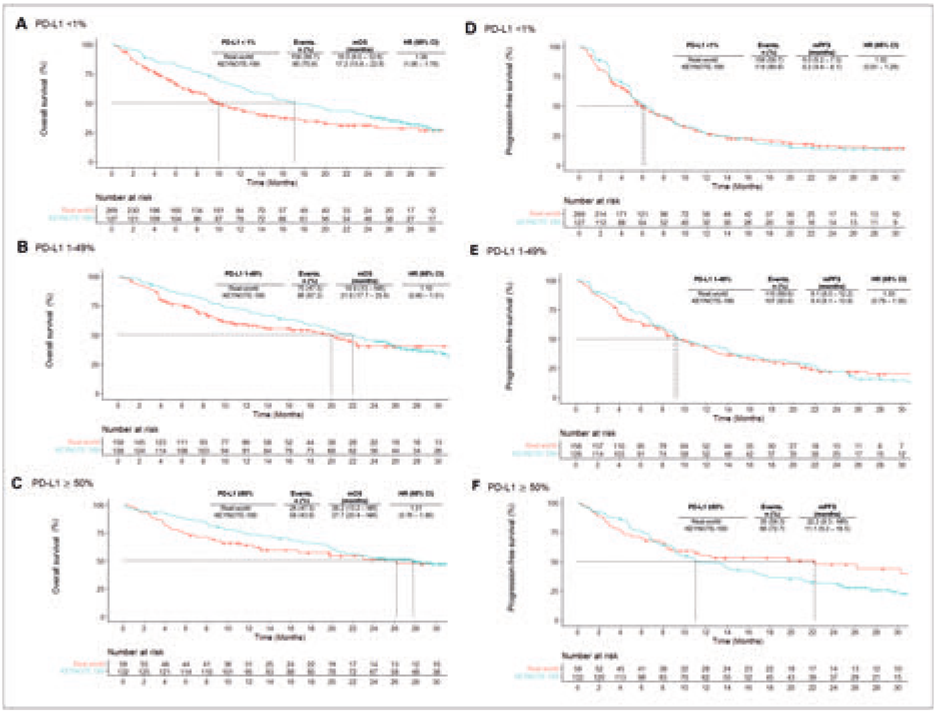


**Figure 2.** Kaplan-Meier survival curves for non-squamous patients stratified for PD-L1 strata: comparison of real-world versus KEYNOTE-189 on overall survival A) PD-L1 <1% stratum B) PD-L1-49% stratum C) PD-L1 >50% stratum and progression-free survival D) PD-L1 <1% stratum E) PD-L1-49% stratum F) PD-L1 >50% stratum.

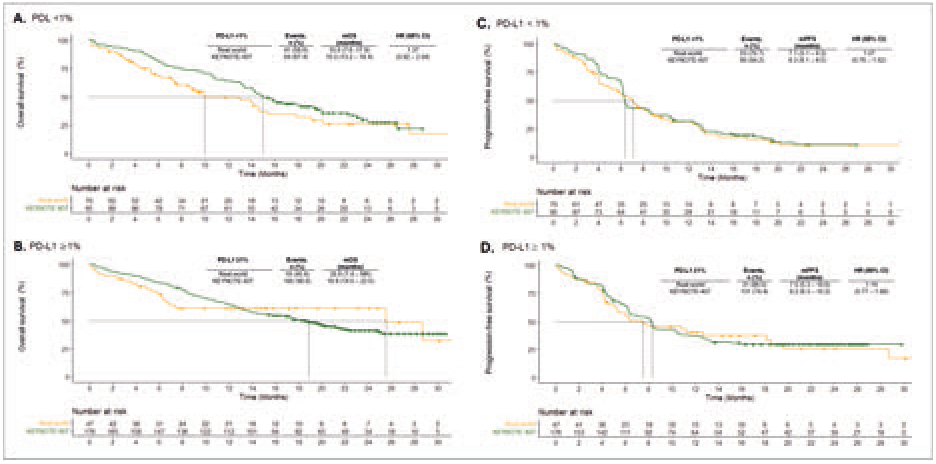


**Figure 3.** Kaplan-Meier survival curves for squamous patients stratified for PD-L1 strata: comparison of real-world versus KEYNOTE-407 on overall survival A) PD-L1 <1% stratum B) PD-L1 ≥1 % stratum and on progression-free survival C) PD-L1 <1% stratum D) PD-L1 ≥1 % stratum.

## Discussion

This study found that in real-world clinical practice, patients with non-squamous metastatic NSCLC and <1% PD-L1 expression who were treated with first line pembrolizumab plus chemotherapy experienced significantly shorter OS compared to the corresponding clinical trial. Additionally, HRs above 1 were observed for OS in the other non-squamous and squamous PD-L1 strata, although these results did not reach statistical significance. For PFS no EE-gap was detected in any of the PD-L1 strata.

Our observation, which demonstrates a tendency towards shorter OS in clinical practice than in the clinical trials, builds upon existing literature on chemotherapy treatment outcome disparities in NSCLC [8] and aligns with previous real-world studies evaluating the effectiveness of pembrolizumab plus chemotherapy in non-squamous metastatic NSCLC patients. For example, the observed one-year survival rates approximate the findings from Liu et al. [13], and Velcheti et al. [14] except for patients with <1% PD-L1 where we observed worse survival rates (44.8% vs 54.5% [13] and 54% [14]).

Explanations for the observed divergence in OS could be differences in the patient and treatment characteristics between the real-world and trial cohorts. In the KEYNOTE-189 and -407 trial, only patients with an ECOG PS of 0 and 1 were enrolled, whereas our study also has patients with an ECOG PS ≥ 2, a significant predictor of poorer survival outcomes [15], as well as patients with unknown ECOG PS. However, after excluding these patients, the EE gap (presented as HR estimate) remained above 1 in all PD-L1 strata indicating less effect even in ECOG PS 0-1 patients (supplementary Table S1). Although it cannot be excluded that in routine care ECOG PS is scored favorably sometimes [16], this suggests that also factors other than the ECOG PS play a role. For example, in the real-world population, less patients received subsequent lines of treatment, and a lower proportion of patients completed four or more cycles of platinum-based chemotherapy during the induction phase. This could indicate differences in patient's perference or physician's awareness of additional treatment options (e.g. investigational products). However, other unmeasured factors are likely to be involved as well. Recently, Lasiter et al. showed that even with a common protocol approach in five datasets, heterogeneity in real-world OS outcomes from pembrolizumab plus chemotherapy persists [17].

Notably, in contrast to OS, no EE gap for PFS was observed in the present study. This, however, does not necessarily mean that there is no gap because the PFS in the real-world could be overestimated. In our previous study investigating the EE gap of durvalumab [18], we observed longer PFS in clinical settings than in the trial. This discrepancy may be attributed to variances in the assessment of disease progression. In clinical trial settings, treatment response is assessed using strict imaging schemes, and independent radiologists interpret these images while adhering strictly to (i)RECIST criteria. On the other hand, in real-world settings, treatment response is assessed with varying frequency of imaging and involves interpretation by different radiologists and thoracic oncologists, who may not consistently adhere to (i)RECIST criteria [19].

A clinically relevant finding of this study is that the OS EE gap was most pronounced in non-squamous patients with <1% PD-L1 expression. Various factors could explain this observation. First, the shorter OS in patients with <1% PD-L1 could be attributable to post progression treatment effects. Notably, only 16.7% of PD-L1<1% patients received subsequent treatment in our study whereas 44.6% of the total trial population received a subsequent treatment (Table 2). Although a lower frequency of subsequent treatments is also observed in the PD-L1 1-49% and ≥50% strata in our study, the impact on survival might be larger in PD-L1 <1% patients. This is supported by the 4-year update of KEYNOTE-189, that showed that second-line treatments were more effective than first-line treatment in PD-L1 <1% patients as measured by PFS, but not for the other PD-L1 strata [20]. Another possible explanation could be that in clinical practice, early discontinuation of immunotherapy, along with chemotherapy, took place more frequently in patients with <1% PD-L1 (82% versus 64% (PD-L1 1-49%) and 75% (PD-L1≥50%), p = 0.04, see Supplementary Table S2). This could suggest a lower conviction to maintain immunotherapy in patients with PD-L1 <1%. At the same time, the recent results of the EMPOWER-3 trial might have contributed decisions to discontinue pembrolizumab, since the trial demonstrated no OS benefit (HR = 1.01, 95%CI 0.63 – 1.60) for patients with PD-L1 <1% treated with cemiplimab plus chemotherapy versus chemotherapy alone [21]. Another possible explanation for the more pronounced EE gap in the PD-L1 <1% stratum could be that fewer patients in clinical practice completed the four cycles of induction chemotherapy. Cytotoxic chemotherapy is considered to enhance modulation of the immune response through PD-L1 inhibition [22][23] and it is conceivable that the non-completion of four cycles has a larger impact in patients with a priori no PD-L1 expression compared to patients with PD-L1 expression at baseline. Lastly, considering the lack of efficacy observed in other immuno-chemotherapy trials, it is possible that the effects seen in patients with <1% PD-L1 treated with pembrolizumab combination in the KEYNOTE-189 trial might be overestimated due to favorable patient characteristics occurring by chance. Unfortunately, aggregated baseline and treatment characteristics nor individual patient level data per PD-L1 stratum from the KEYNOTE-189 are available in the public domain. Access to individual patient level data from trial participants would allow a more comprehensive multi-variable regression analysis into potential factors responsible for the observed EE gap. [24]

In all PD-L1 strata, the real-world mOS for patients treated with pembrolizumab plus chemotherapy was longer compared to the real-world mOS reported by Cramer et.al. for chemotherapy alone in the pre-immunotherapy era, with the smallest survival benefit observed in patients with <1% PD-L1 expression. In this stratum, the absolute mOS difference observed in the real-world was smaller (10.0 versus 8.9 and 6.5 months for cisplatin-pemetrexed and carboplatin-pemetrexed) than in the KEYNOTE-189 trial (17.2 versus 10.2 months). This supports that the real-world effectiveness of pembrolizumab plus chemotherapy is impaired in patients with no PD-L1 tumor expression. Another licensed regimen with nivolumab plus ipilimumab and chemotherapy might be an alternative for these patients [25] because this combination targets both the PD-L1 and CTLA-4 receptors. Unfortunately, real-world comparative effectiveness data of these two regimens in PD-L1 negative patients are not available yet, but needed to judge about the optimal first-line systemic treatment in these specific patients.

The study has several strengths, including its multicenter design, large total sample size and long follow-up time. However, this study also has some limitations. First, both trials lacked data about patient's characteristics per PD-L1 stratum, preventing an evaluation of patient characteristics as explanation for the EE gap. Additionally, the sample sizes of the squamous PD-L1 ≥1% and non-squamous PD-L1≥50% strata are relatively small. While the survival curves showed divergence, there results should be interpreted with caution. Furthermore, the absence of information on the frequency of radiological imaging to determine disease progression limited the investigation of a less strictly timed imaging scheme as a possible explanation for the absence of an EE gap in PFS. Finally, the real-world cohorts had a higher proportion of PD-L1 negative patients compared to the trials (non-squamous cohort: 59.8% versus 31.2% and squamous cohort: 55.4% versus 31.0%). This was partly due to the exclusion of patients treated with pembrolizumab monotherapy for PD-L1 expression ≥50%. Nonetheless, the ratio of PD-L1<1%/PD-L1 1-49% indicates a relatively high proportion of PD-L1 negative patients. This could indicate differences in how PD-L1 status is assessed in the Netherlands. On the other hand, when patients would have been misclassified as no PD-L1 expression, exclusion of misclassified patients would only increase the magnitude of the EE gap in this stratum even further.

## Conclusion

No significant efficacy-effectiveness gap was found for pembrolizumab plus chemotherapy, except in the stratum non-squamous metastatic NSCLC with <1% PD-L1 tumor expression. In these patients, the survival in real-world was considerably shorter compared to the clinical trial results. Further studies are needed to determine which patient, treatment and or context factors contribute to this disparity.

## Funding

This study was supported by Roche Nederland BV.

## References

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin*. 2021;71(3):209-249. doi:10.3322/caac.21660
2. Liu SYM, Zheng MM, Pan Y, Liu SY, Li Y, Wu YL. Emerging evidence and treatment paradigm of non-small cell lung cancer. *J Hematol Oncol*. 2023;16(1):40. doi:10.1186/s13045-023-01436-2
3. Gandhi L, Rodríguez-Abreu D, Gadgeel S, et al. Pembrolizumab plus Chemotherapy in Metastatic Non–Small-Cell Lung Cancer. *N Engl J Med*. 2018;378(22):2078-2092. doi:10.1056/nejmoa1801005
4. Paz-Ares L, Luft A, Vicente D, et al. Pembrolizumab plus Chemotherapy for Squamous Non–Small-Cell Lung Cancer. *N Engl J Med*. 2018;379(21):2040-2051. doi:10.1056/nejmoa1810865
5. Mok TSK, Wu YL, Kudaba I, et al. Pembrolizumab versus chemotherapy for previously untreated, PD-L1-expressing, locally advanced or metastatic non-small-cell lung cancer (KEYNOTE-042): a randomised, open-label, controlled, phase 3 trial. *Lancet*. 2019;393(10183):1819-1830. doi:10.1016/S0140-6736(18)32409-7
6. Hendriks LE, Kerr KM, Menis J, 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. doi:10.1016/j.annonc.2022.12.013
7. Cramer-van der Welle CM, Verschueren M V., Tonn M, et al. Real-world outcomes versus clinical trial results of immunotherapy in stage IV non-small cell lung cancer (NSCLC) in the Netherlands. *Sci Rep*. 2021;11(1):1-9. doi:10.1038/s41598-021-85696-3
8. Cramer-Van Der Welle CM, Peters BJM, Schramel FMNH, Klungel OH, Groen HJM, Van De Garde EMW. Systematic evaluation of the efficacy-effectiveness gap of systemic treatments in metastatic nonsmall cell lung cancer. *Eur Respir J*. 2018;52(6). doi:10.1183/13993003.01100-2018
9. Van De Garde EMW, Plouvier BC, Fleuren HWHA, et al. Pharmacotherapy within a learning healthcare system: rationale for the Dutch Santeon Farmadatabase. *Eur J Hosp Pharm*. 2019;26(1):46-50. doi:10.1136/EJHPHARM-2017-001329
10. Guyot P, Ades AE, Ouwens MJNM, Welton NJ. Enhanced secondary analysis of survival data: Reconstructing the data from published Kaplan-Meier survival curves. *BMC Med Res Methodol*. 2012;12. doi:10.1186/1471-2288-12-9
11. Rodríguez-Abreu D, Powell SF, Hochmair MJ, et al. Pemetrexed plus platinum with or without pembrolizumab in patients with previously untreated metastatic nonsquamous NSCLC: protocol-specified final analysis from KEYNOTE-189. *Ann Oncol*. 2021;32(7):881-895. doi:10.1016/j.annonc.2021.04.008
12. Paz-Ares L, Vicente D, Tafreshi A, et al. A Randomized, Placebo-Controlled Trial of Pembrolizumab Plus Chemotherapy in Patients With Metastatic Squamous NSCLC: Protocol-Specified Final Analysis of KEYNOTE-407. *J Thorac Oncol*. 2020;15(10):1657-1669. doi:10.1016/j.jtho.2020.06.015
13. Liu S V., Hu X, Li Y, Zhao B, Burke T, Velcheti V. Pembrolizumab-combination therapy for previously untreated metastatic nonsquamous NSCLC: Real-world outcomes at US oncology practices. *Front Oncol*. 2022;12(October):1-13. doi:10.3389/fonc.2022.999343
14. Velcheti V, Hu X, Piperdi B, Burke T. Real-world outcomes of first-line pembrolizumab plus pemetrexed-carboplatin for metastatic nonsquamous NSCLC at US oncology practices. *Sci Rep*. 2021;11(1). doi:10.1038/S41598-021-88453-8
15. Gridelli C, Ardizzoni A, Le Chevalier T, et al. Treatment of advanced non-small-cell lung cancer patients with ECOG performance status 2: Results of an European Experts Panel. *Ann Oncol*. 2004;15(3):419-426. doi:10.1093/annonc/mdh087

16. Neeman E, Gresham G, Ovasapians N, et al. Comparing Physician and Nurse Eastern Cooperative Oncology Group Performance Status (ECOG-PS) Ratings as Predictors of Clinical Outcomes in Patients with Cancer. *Oncologist*. 2019;24(12):e1460-e1466. doi:10.1634/theoncologist.2018-0882
17. Lasiter L, Tymejczyk O, Garrett-Mayer E, et al. Real-world Overall Survival Using Oncology Electronic Health Record Data: Friends of Cancer Research Pilot. *Clin Pharmacol Ther*. 2022;111(2):444-454. doi:10.1002/cpt.2443
18. Verschueren M V, Cg A. Durvalumab after chemoradiotherapy in patients with stage III non-small-cell lung cancer : real-world outcomes versus clinical trial results. 2023. doi:10.2217/imt-2023-0002
19. Panageas KS, Ben-Porat L, Dickler MN, Chapman PB, Schrag D. When you look matters: The effect of assessment schedule on progression-free survival. *J Natl Cancer Inst*. 2007;99(6):428-432. doi:10.1093/jnci/djk091
20. Gray J, Rodríguez-Abreu D, Powell SF, et al. FP13.02 Pembrolizumab + Pemetrexed-Platinum vs Pemetrexed-Platinum for Metastatic NSCLC: 4-Year Follow-up From KEYNOTE-189. *J Thorac Oncol*. 2021;16(3):S224. doi:10.1016/j.jtho.2021.01.141
21. Gogishvili M, Melkadze T, Makharadze T, et al. Cemiplimab plus chemotherapy versus chemotherapy alone in non-small cell lung cancer: a randomized, controlled, double-blind phase 3 trial. *Nat Med*. 2022;28(11):2374-2380. doi:10.1038/s41591-022-01977-y
22. Lamberti G, Sisi M, Andrini E, et al. The mechanisms of PD-L1 regulation in non-small-cell lung cancer (NSCLC): Which are the involved players? *Cancers (Basel)*. 2020;12(11):1-21. doi:10.3390/cancers12113129
23. Heenatigala Palliyage G, Samart P, Bobbala S, et al. Chemotherapy-induced PDL-1 expression in cancer-associated fibroblasts promotes chemoresistance in NSCLC. *Lung Cancer*. 2023;181(April):107258. doi:10.1016/j.lungcan.2023.107258
24. Ismail RK, Schramel FMNH, van Dartel M, et al. Individual patient data to allow a more elaborated comparison of trial results with real-world outcomes from first-line immunotherapy in NSCLC. *BMC Med Res Methodol*. 2023;23(1):1-9. doi:10.1186/s12874-022-01760-0
25. Paz-Ares L, Ciuleanu TE, Cobo M, et al. First-line nivolumab plus ipilimumab combined with two cycles of chemotherapy in patients with non-small-cell lung cancer (CheckMate 9LA): an international, randomised, open-label, phase 3 trial. *Lancet Oncol*. 2021;22(2):198-211. doi:10.1016/S1470-2045(20)30641-0

# Supplementary

**Table S1.** Sensitivity analysis for the EE gap for OS and PFS (expressed as HR) for each PD-L1 stratum, only including patients with an ECOG PS 0 and ECOG PS 1.

| **Stratum** | **Number of pembrolizumab patients** | | **HR for OS (95%CI) (real-world vs trial)** | **HR for PFS (95%CI) (real-world vs trial)** |
|---|---|---|---|---|
| | **Real-world (n)** | **Clinical trial (n)** | | |
| Non-squamous cohort | | | | |
| PD-L1 < 1% | 184 | 127 | 1.24 (0.60 – 1.08) | 1.06 (0.77 – 1.28) |
| PD-L1 1-49% | 113 | 128 | 1.17 (0.60 – 1.20) | 1.13 (0.66 – 1.18) |
| PD-L1 ≥ 50% | 39 | 132 | 1.13 (0.52 - 1.51) | Non proportional hazard |
| Squamous cohort | | | | |
| PD-L1 < 1% | 45 | 95 | 1.32 (0.49 – 1.19) | 1.14 (0.60 – 1.30) |
| PD-L1 ≥ 1% | 26 | 176 | Non-proportional hazard | 0.91 (0.68 – 1.76) |

**Table S2.** Patients who prematurely discontinue platinum-based chemotherapy (<4 cycles) and also prematurely discontinue pembrolizumab (<4 cycles), stratified per PD-L1 stratum.

| | **PD-L1 <1% (n = 269)** | **PD-L1 1-49% (n = 158)** | **PD-L1 ≥ 50% (n =59)** | **p value** |
|---|---|---|---|---|
| Premature discontinuation (<4 cycles) of chemotherapy (n, (%)) | 99 (36.7) | 55 (34.8) | 16 (27.1) | 0.37 |
| Premature discontinuation (< 4 cycles) of pembrolizumab in patients who prematurely discontinue chemotherapy (n, (%*)) | 81 (81.8) | 35 (63.6) | 12 (75.0) | 0.04 |

*percentage was calculated from the number of patient who prematurely discontinue chemotherapy .

# 2.3

# Comparative effectiveness of chemotherapy with nivolumab plus ipilimumab versus chemotherapy with pembrolizumab in PD-L1 negative advanced stage NSCLC patients.

Marjon V Verschueren, Dagmar TA Hiensch, Peter MJ Plomp,
Lisanne A Kastelijn, Ewoudt MW van de Garde, Bas J M Peters

*Clinical Lung Cancer. 2025 Jan; 26(4):243-248*
*DOI: 10.1016/j.cllc.2025.01.009*

# Abstract

## Background

Recently, the combination of nivolumab, ipilimumab and chemotherapy (NIC) became available for the treatment of metastatic non-small cell lung cancer (mNSCLC). This study aimed to compare the real-world outcomes of NIC with the current standard of care pembrolizumab plus chemotherapy (PC) in PD-L1 negative mNSCLC patients treated in clinical practice and to compare these outcomes with the results of the Checkmate-9LA trial.

## Methods

mNSCLC patients with PD-L1<1% treated with NIC or PC at two large hospitals in the Netherlands between 2019 and 2024 were included. The objective response rate (ORR) progression-free survival (PFS) and treatment characteristics of patients treated with NIC were compared to those of patients treated with PC. Additionally, the outcomes from NIC were compared to the results from the CheckMate9LA trial. A multivariate Cox regression was used to calculate PFS hazard ratios (HR).

## Results

PD-L1 negative mNSCLC patients treated with NIC had a higher ORR than those treated with PC (41% versus 27%). The PFS was slightly longer for patients treated with NIC versus PC (5.5 versus 4.5 months, aHR=0.91 [95%CI0.59–1.58]), although not statistically significant. Treatment discontinuation rates were comparable between the NIC and PC cohorts (72% vs 68%). The outcomes for patients treated with NIC in clinical practice were comparable to the Checkmate-9LA trial.

## Conclusion

For mNSCLC patients with <1% PD-L1 expression, the treatment responses to NIC were numerically better than to PC. Larger cohorts with longer follow-up and survival outcomes are needed to further establish the role of NIC in PD-L1 negative patients.

## Introduction

In patients with non-squamous metastatic non-small cell lung cancer (mNSCLC) and <1% PD-L1 expression, the combination of pembrolizumab (PD-1 inhibitor) plus chemotherapy (PC) has been the standard first-line treatment, based on the KEYNOTE-189 trial [1,2]. Recently, the Checkmate-9LA trial introduced a new regimen combining nivolumab (PD-1 inhibitor) and ipilimumab (CTLA-4 inhibitor) with chemotherapy (NIC), showing a survival benefit for patients with <1% PD-L1 expression, who are under 75, have a smoking history, and lack liver metastases [3]. As a result, NIC is now offered in the Netherlands to mNSCLC patients fitting these criteria. A network meta-analysis, which indirectly compared the efficacy of these treatments, suggested that NIC was associated with better overall survival (OS) than PC in PD-L1 negative patients [4]. However, a head-to-head trial is lacking, so currently the best treatment option in PD-L1 negative patients remains unknown.

Unfortunately, survival outcomes reported in randomized controlled trials (RCTs) are rarely mirrored in daily practice as real-world patients differ significantly from those in RCTs. Patients in daily practice are often older, have more comorbidities, and a worse performance status leading to worse survival outcome. Indeed, the overall survival (OS) of mNSCLC patients receiving systemic treatments in clinical practice is notably lower than the OS reported in the trials [5,6].

This study aims to assess the real-world outcomes of nivolumab plus ipilimumab combined with chemotherapy in mNSCLC patients with <1% PD-L1 expression and to compare these results with historical controls of mNSCLC <1% PD-L1 patients treated with pembrolizumab and chemotherapy. In addition, these outcomes will be compared to the PD-L1<1% subgroup from the Checkmate-9LA to assess how clinical trial results translate to real-world practice.

## Methods

### Study objective and design

The objective of this multicenter, historical cohort study is to compare the effectiveness of NIC and PC in mNSCLC patients with <1% PD-L1 expression. Patients who received at least one cycle of first-line NIC at the St. Antonius hospital in Nieuwegein/Utrecht or Isala hospital in Zwolle between 1 January 2022 and 1 October 2024were included in the real-world NIC cohort. These patients were compared to historic controls diagnosed with mNSCLC and <1% PD-L1 expression who were treated with PC at St. Antonius hospital between 1 January 2019 and 31 December 2021 (real-world PC cohort) (13). Patients were excluded if actionable genomic alterations such as EGFR and ALK were found. All patients with squamous histology without smoking history and non-squamous histology were tested for actionable genomic alterations., The outcomes of the real-world NIC cohort were compared to those of the PD-L1 <1% patient cohort treated with NIC in the Checkmate9LA trial (CM-9LA) (4). This study followed the observational research guidelines of the European Society for Medical Oncology Guidance for reporting real-world evidence (ESMO GROW) [7].

## Data source and variables

The following baseline characteristics were manually extracted from medical charts 30 days before and 30 days after the start date of the treatment: age, gender, eastern cooperative oncology group performance status (ECOG PS), body mass index (BMI), smoking status, KRAS mutation, and the presence and location of metastases. Follow-up information was extracted until the end of this study on 23 December 2024. All study data were collected, anonymized, and managed in REDCap electronic data capture tool.

## Outcomes

The primary outcome of this study was the objective response rate (ORR), defined as the percentage of patients with a complete or partial response as their best overall response. The study also assessed the disease control rate (DCR), defined as the percentage of patients with a complete or partial response or stable disease as their best overall response. The patient’s response to treatment was determined according to the local protocol, which included semi-standardized response assessment schedules. Radiological imaging with (PET-)CT and/or MRI was scheduled every two to four cycles of treatment, and where possible, the (immune) response evaluation criteria in solid tumors (RECIST) were used. Radiology reports from baseline and follow-up scans were used to extract the diameter of target lesions for each patient. If RECIST criteria were not used, the best overall response reported in the medical notes by the thoracic oncologist was used. In cases where the reported response was unclear, an independent thoracic oncologist reassessed the radiological images. Additionally, the number of chemotherapy cycles administered, the discontinuation of immunotherapy (including reasons for discontinuation), and the duration of treatment - defined as the time in months from treatment initiation to the last administration of immunotherapy - were evaluated. Lastly, PFS was evaluated by calculating the time between the date of treatment initiation and the date of disease progression, death, or censoring, whichever occurred first. Patients were censored on the last date of clinic visit before the end of follow-up (27 June 2024).

## Statistical analysis

Statistical analysis was done withR version 4.1.2.. The baseline characteristics and treatment outcomes of the real-world NIC and PC cohorts were compared using appropriate statistical tests. For each patient, the treatment response was visualized in a vertical bar plot by calculating the change of the sum of the largest diameters of the target lesion between the baseline and follow-up scans. Additionally, the median PFS (mPFS) was calculated, and Hazard Ratios (HRs) with 95% confidence intervals (CIs) for PFS were calculated using both univariable and multivariable Cox regression models. The multivariable Cox regression model was constructed with the type of treatment regimen, ECOG PS, and all other variables with a p-value of <0.2 from the univariable Cox-regression analysis. Lastly, a sensitivity analysis was performed to investigate the ORR, DCR and best overall response rate after excluding all patients with an ECOG PS ≥ 2.

### Ethical statement

The study was approved by the medical-ethical review committee (Z23.088 (St. Antonius hospital) and 20231110 (Isala hospital)). Because of the retrospective methodology and anonymization of the data no informed consent was asked. The study was conducted in compliance with the 1964 Helsinki Declaration and its later amendments.

## Results

A total of 58 patients were included in the NIC cohort and 60 patients were included in the PC cohort. In total, 361 patients were included in the NIC arm of the CM-9LA trial, of which 135 patients were PD-L1 negative.

### Baseline characteristics

Table 1 presents the patient and disease characteristics of the real-world and trial cohorts. The proportion of real-world patients with ECOG PS ≥ 2 was smaller for patients treated with NIC than for patients treated with PC (7% vs 35%, p <0.05). Other characteristics were not statistically different between both real-world cohorts. The median follow-up time was 7.4 months (Interquartile range [IQR]: 4.7-9.6) for patients treated with NIC and 6.1 months (IQR: 3.2-11.9) for those treated with PC. NIC patients in the real world had more ECOG PS ≥ 2 (7% vs 0%) and bone metastases (46% vs 27%) than CM-9LA participants.

**Table 1.** Patient and disease characteristics of mNSCLC patients with <1% PD-L1 expression included in the real-world NIC cohort and real-world PC cohort and the characteristics from the total patient population treated with nivolumab and ipilumumab plus chemotherapy in the Checkmate 9LA trial cohort.

| | **Real world NIC (n=58)** | **Real world PC (n=60)** | ***p* value** | **CM-9LA all PD-L1 cohorts (n=361)** |
|---|---|---|---|---|
| Age, years | | | | |
| Median (IQR) | 67 (60-71) | 66 (59-73) | 0.87 | 64 (59-70) |
| Gender (n(%)) | | | | |
| Male | 31 (53) | 36 (60) | 0.57 | 252 (70) |

**Table 1.** Continued.

| | Real world NIC (n=58) | Real world PC (n=60) | *p* value | CM-9LA all PD-L1 cohorts (n=361) |
|---|---|---|---|---|
| ECOG PS (n(%)) | | | | |
| 0-1 | 51 (88) | 39 (65) | | 361 (100) |
| ≥2 | 4 (7) | 21 (35) | < 0.05 | 0 |
| Unknown | 3 (22) | 0 | | - |
| BMI, kg/m$^2$ (n(%)) | | | | |
| Mean (SD) | 25.1 (4.5) | 25.8 (4.3) | 0.35 | - |
| Smoking status (n(%)) | | | | |
| Never | 1 (2) | - | | 46 (13) |
| Former or current | 56 (97) | - | - | 315 (87) |
| Unknown | 1 (2) | 60 (100) | | - |
| Histology (n(%)) | | | | |
| Squamous | 10 (17) | 12 (20) | | 113 (31) |
| Non-squamous | 47 (81) | 48 (80) | 0.81 | 248 (69) |
| Other | 1 (2) | 0 | | 0 |
| KRAS (n (%) | | | | |
| Wildtype | 37 (64) | 28 (47) | | - |
| Mutation | 14 (24) | 21 (35) | 0.10 | - |
| Not tested | 7 (12) | 11 (18) | | - |
| Metastases (n(%)) | | | | |
| Bone | 23 (40) | 20 (33) | 0.60 | 97 (27) |
| Brain | 5 (9) | 5 (8) | 0.94 | 64 (18) |
| Liver | 5 (9%) | 7 (12%) | 0.72 | 68 (19) |
| Follow-up time, months | | | | |
| Median (IQR) | 7.6 (4.7 – 9.6--) | 6.1 (3.2-11.9) | 0.79 | 12.2 |

ECOG PS = Eastern Cooperative Oncology Group Performance Status, BMI = Body Mass Index.

## Response

Figure 1 presents the change in sum of the largest diameters of the target lesions per patient, and Table 2 presents the overall treatment outcomes. The ORR was numerically, but not statistically significant higher in patients treated with NIC than those treated with PC (41% vs 27%, p = 0.09). Similarly, the DCR was higher in the real-world NIC cohort than in the PC cohort, but also not statistically significant (74% vs 67%, p = 0.49). In the NIC cohort, the most common best overall response was partial response (39%). In the PC cohort, the most common best overall response was stable disease (40%). The treatment outcomes were comparable after excluding patients with an ECOG PS ≥ 2 (ORR [41% vs 23%, p = 0.16] and DCR [74% vs 72%, p = 0.83] (Table S1). The ORR in the real-world NIC cohort was higher than in the 9LA trial cohort (41% vs 31%), whereas the DCR was lower (74% vs 82%).

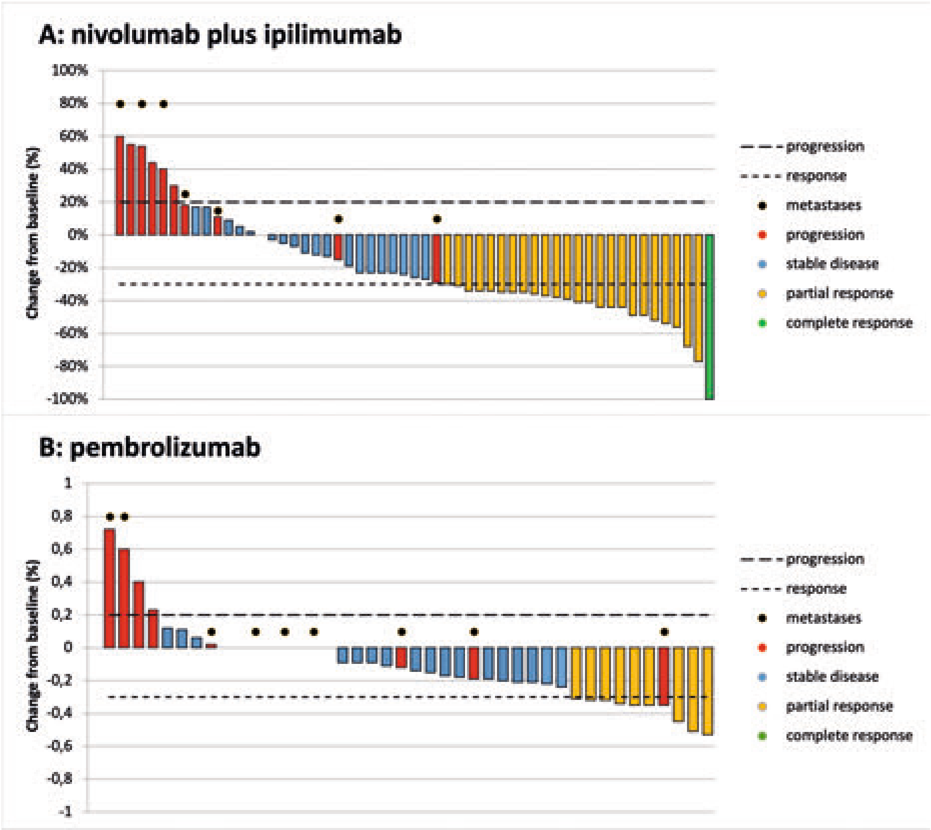


**Figure 1.** Change in sum of the largest diameters of target lesions from baseline in real-world patients treated withNIC, n=52* (A) and CP, n=52*(B). *Response assessments were limited to patients for whom it was possible to calculate response.

**Table 2.** Objective response rate, disease control rate, best overall response for the real-world NIC, real-world PC and Checkmate-9LA PD-L1 <1% cohorts.

| | Real world NIC (n=58) | Real world PC (n=60) | *p* value | CM-9LA PD-L1<1% cohort (n=135) |
|---|---|---|---|---|
| Objective response rate | 24 (41) | 16 (27) | 0.09 | 42 (31) |
| Disease control rate | 43 (74) | 40 (67) | 0.49 | 110 (82) |
| Best overall response (n(%)) | | | | |
| *Complete response* | 1 (2) | 0 | | 2 (2) |
| *Partial response* | 23 (39) | 16 (27) | | 40 (30) |
| *Stable disease* | 19 (33) | 24 (40) | - | 68 (50) |
| *Progressive disease* | 10 (17) | 13 (22) | | 11 (8) |
| *Unknown* | 5 (9) | 7 (12) | | 14 (10) |

## Survival outcomes

The mPFS for the real-world NIC cohort was 5.5 months (95% CI 4.0- 8.9) and for the PC cohort 4.5 months (95% CI 2.6-6.5) with a non-significant aHR of 0.91 [95% CI 0.59-1.58] p=0.89. In the multivariable model for PFS, ECOG PS ≥ 2 (vs ECOG 0-1) and the presence of brain, bone or liver metastases (vs none of these metastases), were statistically significant associated with worse PFS outcomes, with aHR 2.10 (95%CI 1.21 - 3.64) and aHR 2.04 (95%CI 1.29 -3.25), respectively (Table S2). In the CM-9LA a mPFS of 5.8 months (95%CI 4.4-7.6) was reported in the PD-L1 negative subgroup treated with NIC.

## Treatment exposure and discontinuation

Table 3 presents the treatment exposure and discontinuation for the real-world cohorts and trial cohort.

The median duration of treatment was comparable between real-world patients treated with NIC and treated with PC (3.6 vs 3.5 months, respectively). In both the real-world NIC and real-world PC cohort, most patients discontinued treatment (72% and 78%, respectively), with disease progression being the most common reason for discontinuation (67% and 64%, respectively). The median duration of treatment for patients treated in the CM-9LA trial cohort was longer than for patients treated in the real-world NIC cohort (6.1 vs 3.6 months). The other treatment characteristics were comparable between the trial and the real-world NIC cohort.

**Table 3.** Treatment exposure and discontinuation for the patients treated in the real-world NIC- and PC- cohorts and for all patients treated in CM-9LA trial.

| | | **Real world NIC (n=58)** | **Real world PC (n=60)** | ***p-value*** | **CM-9LA total population[*] (n = 358)** |
|---|---|---|---|---|---|
| Duration of treatment, median (IQR) | | 3.6 [2.1 – 5.6] | 3.5 [0.7 – 6.7] | 0.30 | 6.1 [2.7-13.5] |
| Number of chemotherapy cycles[$], n(%) | 0 | 1 (2) | 0 (0) | 0.12 | 0 (0) |
| | 1 | 5 (9) | 9 (15) | | 25 (7) |
| | ≥2 | 52 (89) | 51 (85) | | 333 (93) |
| Discontinuation immunotherapy[&], n(%) | | 42 (72) | 47 (78) | 0.30 | 281 (79) |
| Reason, n/N (%): | Progressive disease | 28 (67) | 30 (64) | 0.70 | 175 (62) |
| Toxicity | | 8 (19) | 6 (13) | | 65 (23) |
| Death | | 2 (5) | 4 (9) | | 3 (1) |
| Other | | 4 (10) | 7 (15) | | 38 (14) |

[*] The CM-9LA trial did not report these findings separately for patients with PD-L1 <1%.

[$] The standard treatment for NIC in real-world and CM-9LA trial settings consists of 2 cycles of chemotherapy, and the standard treatment for PC in real-world setting consists of 4 cycles of chemotherapy.

[&] Patients who discontinued immunotherapy because of treatment completion were not included (n = 0 for NIC and n = 4 for PC).

## Discussion

In this cohort study we found that the ORR in PD-L1 negative mNSCLC patients treated with NIC was higher than in those treated with PC, although not statistically significant. Additionally, the point estimate for PFS was marginally longer for patients treated with NIC than for those treated with PC, although also not statistically significant. Furthermore, the treatment discontinuation rates were comparable between both real-world cohorts. Lastly, the outcomes for real-world patients treated with NIC were similar to those reported in the CheckMate-9LA trial [3]. Our early assessment offers valuable insights into the initial treatment response to NIC in clinical practice. These early effectiveness findings suggest treatment with NIC over PC in patients with <1% PD-L1 expression. However, a larger cohort and longer follow-up towards OS are needed to confirm this conclusion.

Our findings build upon existing comparative effectiveness studies in this population. Matsumoto et al. showed in their matched cohort study a longer median OS (mOS) in PD-L1 negative patients treated with NIC than in those treated with PC (26.0 versus 17.0 months; HR 0.49) [8]. Their mOS observed for NIC exceeds the 17.7 months reported in the CM-9LA significantly. This is unexpected, as clinical practice outcomes are generally worse than those seen in RCTs. A second observational study also reported a favourable mOS for NIC compared to PC (16.0 versus 12.0 months) for the PD-L1 negative subgroup [9].

We also compared our findings of real-world PD-L1 negative patients treated with NIC with the outcomes of those treated with NIC in the CM-9LA. In our study, the observed ORR (41%) was numerically higher, and the mPFS (5.5 months) was similar to the findings from the CM-9LA trial (31% and 5.8 months, respectively). This suggests that patients treated in clinical practice may have at least a similar treatment response to those in trial settings despite having worse performance scores. However, comparisons of treatment response and PFS outcomes should be interpreted with caution due to potential variability in their assessment. In clinical practice, treatment responses are measured less frequently, and the (immune) RECIST criteria are often not used, unlike the standardized methods in clinical trials. Therefore, although our early effectiveness findings are encouraging, further real-world studies with OS endpoints are needed to confirm these results.

The main limitations of this study were its small sample size and relatively short follow-up. As this study should be considered as an initial early assessment, more patients with longer follow-up are needed in future research to increase statistical power and enable the evaluation of long-term outcomes such as OS. At present, a Dutch nationwide cohort study is being conducted that is designed to evaluate the OS of PD-L1 negative patients treated with NIC. Another limitation is that we did not collect information on the assessment schedules and criteria used to measure treatment response, which hindered us from assessing the impact of any discrepancies in these measurements on our findings. However, since overall duration of this cohort study was only two years, we do not expect significant differences in the way treatment response was measured between patients treated with NIC and PC.

In conclusion, a non-significant better treatment response was observed for mNSCLC patients with <1% PD-L1 expression treated with NIC than those treated with PC, suggesting a potential benefit of NIC in this population. The treatment discontinuation rates between both real-world cohorts were comparable. Additionally, our findings of patients treated with NIC were comparable to the CM-9LA. Observational studies with more patients, longer follow-up and OS outcomes are needed to determine the optimal therapeutic approach for patients with mNSCLC and low PD-L1 expression

## References

1. S. Novello, D.M. Kowalski, A. Luft, M. Gümüş, D. Vicente, J. Mazières, J. Rodríguez-Cid, A. Tafreshi, Y. Cheng, K.H. Lee, A. Golf, S. Sugawara, A.G. Robinson, B. Halmos, E. Jensen, P. Schwarzenberger, M.C. Pietanza, L. Paz-Ares, Pembrolizumab Plus Chemotherapy in Squamous Non-Small-Cell Lung Cancer: 5-Year Update of the Phase III KEYNOTE-407 Study., J. Clin. Oncol. 41 (2023) 1999–2006. https://doi.org/10.1200/JCO.22.01990.
2. M.C. Garassino, S. Gadgeel, G. Speranza, E. Felip, E. Esteban, M. Dómine, M.J. Hochmair, S.F. Powell, H.G. Bischoff, N. Peled, F. Grossi, R.R. Jennens, M. Reck, R. Hui, E.B. Garon, T. Kurata, J.E. Gray, P. Schwarzenberger, E. Jensen, M.C. Pietanza, D. Rodríguez-Abreu, Pembrolizumab Plus Pemetrexed and Platinum in Nonsquamous Non-Small-Cell Lung Cancer: 5-Year Outcomes From the Phase 3 KEYNOTE-189 Study., J. Clin. Oncol. 41 (2023) 1992–1998. https://doi.org/10.1200/JCO.22.01989.
3. D.P. Carbone, T.-E. Ciuleanu, M. Schenker, M. Cobo, S. Bordenave, O. Juan-Vidal, J. Menezes, N. Reinmuth, E. Richardet, Y. Cheng, H. Mizutani, E. Felip, B. Zurawski, A. Alexandru, L. Paz-Ares, S. Lu, T. John, X. Zhang, J. Mahmood, N. Hu, T. De, I. Santi, J.R. Penrod, Y. Yuan, A. Lee, M. Reck, Four-year clinical update and treatment switching-adjusted outcomes with first-line nivolumab plus ipilimumab with chemotherapy for metastatic non-small cell lung cancer in the CheckMate 9LA randomized trial., J. Immunother. Cancer 12 (2024). https://doi.org/10.1136/jitc-2023-008189.
4. L. Liu, H. Bai, C. Wang, S. Seery, Z. Wang, J. Duan, S. Li, P. Xue, G. Wang, Y. Sun, X. Du, X. Zhang, Z. Ma, J. Wang, Efficacy and Safety of First-Line Immunotherapy Combinations for Advanced NSCLC: A Systematic Review and Network Meta-Analysis., J. Thorac. Oncol. 16 (2021) 1099–1117. https://doi.org/10.1016/j.jtho.2021.03.016.
5. M. V Verschueren, B.J. Peters, L.T. Bloem, V.R. Kruik, E.B. Uitvlugt, A.R. Bijsmans, A.C. Egberts, E.M. van de Garde, Pembrolizumab Plus Chemotherapy Per PD-L1 Stratum In Patients With Metastatic Non-Small Cell Lung Cancer: Real-World Effectiveness Versus Trial Efficacy., Clin. Lung Cancer 25 (2024) 119-127.e1. https://doi.org/10.1016/j.cllc.2023.12.011.
6. C.M. Cramer-van der Welle, B.J.M. Peters, F.M.N.H. Schramel, O.H. Klungel, H.J.M. Groen, E.M.W. van de Garde, Santeon NSCLC Study Group, Santeon NSCLC study group, Systematic evaluation of the efficacy-effectiveness gap of systemic treatments in metastatic nonsmall cell lung cancer., Eur. Respir. J. 52 (2018). https://doi.org/10.1183/13993003.01100-2018.
7. Castelo-Branco L, Pellat A, Martins-Branco D, Valachis A, Derksen JWG, Suijkerbuijk KPM, et al. ESMO Guidance for Reporting Oncology real-World evidence (GROW). ESMO Real World Data and Digital Oncology. 2023 Nov;1:100003.
8. K. Matsumoto, T. Shiroyama, M. Tamiya, T. Minami, Y. Kinehara, A. Tamiya, Y. Suga, T. Kuge, M. Mori, H. Suzuki, S. Tobita, K. Ueno, Y. Namba, S. Tetsumoto, T. Niki, O. Morimura, A. Osa, K. Nishino, I. Nagatomo, Y. Takeda, T. Kijima, A. Kumanogoh, Real-world outcomes of nivolumab plus ipilimumab and pembrolizumab with platinum-based chemotherapy in advanced non-small cell lung cancer: a multicenter retrospective comparative study., Cancer Immunol. Immunother. 73 (2024) 4. https://doi.org/10.1007/s00262-023-03583-4.
9. W. Shalata, N. Maimon Rabinovich, A. Agbarya, A. Yakobson, Y. Dudnik, A. Abu Jama, A.Y. Cohen, S. Shalata, A. Abu Hamed, T. Ilan Ber, O. Machluf, G. Shoham Levin, A. Meirovitz, Efficacy of Pembrolizumab vs. Nivolumab Plus Ipilimumab in Metastatic NSCLC in Relation to PD-L1 and TMB Status., Cancers (Basel). 16 (2024). https://doi.org/10.3390/cancers16101825.

# Supplementary

**Table S1.** Objective response rate, disease control rate and best overall response for the real-world NIC and real-world PC after excluding all patients with ECOG PS ≥ 2.

| | Real-world NIC (n = 54) | Real-world PC (n = 39) | *p value* |
|---|---|---|---|
| Objective response rate | 22 (41) | 9 (23) | 0.16 |
| Disease control rate | 40 (74) | 28 (72) | 0.83 |
| Best overall response (n(%)) | | | |
| *Complete response* | 1 (2) | 0 (0) | |
| *Partial response* | 21 (4) | 9 (23) | |
| *Stable disease* | 18 (33) | 19 (49) | - |
| *Progressive disease* | 10 (19) | 9 (23) | |
| *Unknown* | 4 (7) | 2 (5) | |

Abbreviations; ECOG PS, Eastern Cooperative Oncology Group Performance Score; NIC, nivolumab and ipilimumab plus chemotherapy; PC, pembrolizumab plus chemotherapy

**Table S2. Uni and multivariable cox regression models for progression-free survival**

| Variable | HR for PFS | 95% CI | pvalue | aHR for PFS | 95%CI | *p value* |
|---|---|---|---|---|---|---|
| | Univariable analysis | | | Multivariable analysis | | |
| Type of treatment | | | | | | |
| PC | ref. | 0.58- 1.42 | 0.68 | 0.91 | 0.59 – 1.58 | 0.89 |
| NIC | 0.81 | | | | | |
| Age | | | | | | |
| <65 | ref. | 0.55 – 1.32 | 0.47 | 0.71 | 0.44 – 1.13 | 0.15 |
| ≥65 | 1.18 | | | | | |
| Gender | | | | | | |
| male | ref. | 0.59- 1.44 | 0.73 | | | |
| female | 0.92 | | | | | |
| ECOG | | | | | | |
| PS 0 or 1 | ref. | 1.05 – 2.85 | 0.03 | 2.10 | 1.21 – 3.64 | <0.01 |
| PS ≥2 | 1.73 | | | | | |
| BMI | | | | | | |
| <25 | ref. | 0.44-1.08 | 0.11 | 0.73 | 0.46 – 1.15 | 0.18 |
| ≥25 | 0.69 | | | | | |
| Histology | | | | | | |
| non-squamous | ref. | 0.45-1.43 | 0.45 | | | |
| squamous | 0.80 | | | | | |
| Brain, bone or liver metastasis | | | | | | |
| no | ref. | 1.16- 2.78 | <0.01 | 2.04 | 1.29 – 3.25 | <0.01 |
| yes | 1.80 | | | | | |

Abbreviations: aHR, adjusted Hazard Ratio; BMI, body mass index; ECOG PS, Eastern Cooperative Oncology Group Performance Status; NIC, nivolumab and ipilimumab plus chemotherapy; PC, pembrolizumab plus chemotherapy.

# 2.4

# The association between gut microbiome affecting concomitant medication and the effectiveness of immunotherapy in patients with stage IV NSCLC

Marjon V Verschueren, Christine M Cramer- van der Welle, Merel Tonn,
Franz MNH Schramel, Bas J M Peters, Ewoudt MW van de Garde

*Scientific Reports. 2021 Dec 2;11(1):23331.*
*DOI: 10.1038/s41598-021-02598-0*

# Abstract

## Objectives

Several observational studies suggested that gut microbiome-affecting-medication impairs the effectiveness of immunotherapy in patients with metastatic non-small-cell lung cancer (NSCLC). We postulated that if the effectiveness of immunotherapy is affected by drug-related changes of the microbiome, a stronger association between the use of co-medication and overall survival (OS) will be observed in patients treated with immunotherapy as compared to patients treated with chemotherapy.

## Methods

In a retrospective matched cohort study, immunotherapy patients were matched (1:1) to patients treated with chemotherapy in the pre immunotherapy era. The association between the use of antibiotics, opioids, proton pump inhibitors, metformin and other antidiabetics on OS was assessed with multivariable cox-regression analyses. Interaction tests were applied to investigate whether the association differs between patients treated with immuno- or chemotherapy.

## Results

A total of 442 patients were studied. The use of antibiotics was associated with worse OS (adjusted Hazard Ratio (aHR) 1.39, p=0.02) independent of the type of therapy (chemotherapy or immunotherapy). The use of opioids was also associated with worse OS (aHR 1.33, p=0.01). The other drugs studied showed no association with OS. Interaction term testing showed no effect modification by immuno- or chemotherapy for the association of antibiotics and opioids with OS.

## Conclusion

The use of antibiotics and opioids is similarly associated with worse outcomes in both chemotherapy and immunotherapy treated NSCLC patients. This suggests that the association is likely to be a consequence of confounding rather than disturbing the composition of the microbiome.

## Introduction

Lung cancer is the leading cause of cancer-related deaths worldwide. In 2012, the World Health Organisation (WHO) estimated that of the 1.6 million lung cancer related deaths, non-small-cell lung cancer (NSCLC) is the most frequent histological type, representing 85% of cases [1].

Since the beginning of systemic treatment of stage III and IV NSCLC patients, platinum-doublet chemotherapy has been the first-choice treatment. With the development of new therapies, more effective treatment options have become available in addition to the platinum-doublet therapy [2] [3]. For patients without driver mutations, immunotherapies targeting immune checkpoints, such as the PD-1/PD-L1 pathway, have become an important new treatment option. Currently, four PD-1/PD-L1 inhibitors (pembrolizumab, nivolumab, durvalumab and atezolizumab) have been approved for treatment of patients with advanced NSCLC.

The approval of immunotherapy, with or without chemotherapy, is based on several phase III studies demonstrating superior efficacy in a population meeting the strict in- and exclusion criteria [4] [5] [6]. For example, patients with present active infections, autoimmune conditions, and other comorbidities often requiring concomitant medication are not included in these trials. Therefore, the impact of these factors, including the use of concomitant medication on the efficacy of immunotherapy, cannot be studied from clinical trial data.

Recent studies demonstrate that the microbial flora plays an important role in modulation of immunotherapy effectiveness by affecting the tumor immuno-microenvironment [7]. It is well known that the microbiome can vary significantly from one individual to another, which has been proposed as an explanation of the variability of response to immunotherapy [8]. Routy et al. showed that changes in composition of the gut microbiome negatively influence the outcome of PD-1 inhibition by immunotherapy in mice and patients. Additionally, the study found that the replacement of deleterious microbial flora by a favorable one, can restore the efficacy of the immunotherapy response in mice [9]. The composition of the human microbiome is influenced by several factors such as host genetics, lifestyle factors and the use of antibiotics. Besides antibiotics, recent microbiome studies showed that many other drugs can also alter the composition of the microbiome, as investigated with high-throughput drug screens and metagenomics analyses [10] [11].

Accumulating observational studies show that antibiotic treatment associates significantly with attenuated clinical outcomes in NSCLC patients treated with immunotherapy [9] [12] [13]. However, the majority of these studies analyzed the effects of antibiotics on survival outcomes in cohort studies including only immunotherapy patients. This makes the results susceptible for bias due to confounding by indication, a situation in which patient characteristics - rather than the presence of concomitant medication- are independent predictors of clinical outcomes. Therefore, it is still an active area of debate whether or not there is a causal link between antibiotic related changes of the microbiome and the decreased effectiveness of immunotherapy. Assuming that a prospective clinical trial randomizing based on concomitant medication will not conducted, additional observational studies with alternative designs are needed to resolving this matter, for example a study design including a control group not treated with immunotherapy. If the association is causal one would expect the association to be more pronounced with immunotherapy compared to chemotherapy. The aim of the present study is to test this hypothesis.

## Patients and methods

### Study design

This is a retrospective matched cohort study using clinical data from six hospitals of the Santeon hospital network.

### Data sources

The Santeon hospital network consists of seven large teaching hospitals geographically spread across the Netherlands. The Netherlands Cancer Registry (NCR) was used for identifying all patients diagnosed with NSCLC in a Santeon hospital, and for obtaining information on the date of diagnosis and the vital status. Individual patients are assigned a unique anonymous identifier, which enables them to be tracked in the Santeon Farmadatabase (SFD). The information of the SFD was used for collection of detailed information about the systemic treatments [14]. Finally, the patients' medical records were used to complement the database with detailed information about the clinical and demographic characteristics, the use of concomitant medication and the treatment response. All data were gathered and stored at a Research Electronic Data Capture database (REDcap) [15].

### Study population

Patients diagnosed with stage IV NSCLC between 1st January 2015 to 1st January 2019 and who started first-, second- or third-line immunotherapy before the 1st of January 2020 were assigned to the immunotherapy group. We matched every patient in the immunotherapy group to a patient with stage IV NSCLC who received conventional chemotherapy in the pre-immunotherapy era, and has been diagnosed before 1st January 2015 (see publication of Cramer et al for more details about this cohort) [16].

Patients were (1:1) matched on gender, age groups (<50 year, 50-60 year, 61-70 year, 71-80 year, >80 year) and line of treatment (first, second or third).

### Clinical characteristics

The patient characteristics and demographics were collected manually from the patients' medical records, including age, gender, body mass index (BMI), Eastern Cooperative Oncology Group-Performance status (ECOG-PS), the histology subtype, brain metastases and lines of systemic treatments. First-line treatment (1L) was defined as the initial systemic therapy used in the treatment of NSCLC. Second-line and third-line treatment (2L and 3L) were defined as the therapy given after discontinuation due to disease progression or completion of first or second-line treatment, respectively. Additionally, oncogenic driver mutation status (e.g. EGFR, ALK, ROS-1) and PD-L1 expression were collected if available.

### Concomitant medication

Patient records were reviewed to collect information on the use of concomitant medication potentially affecting the microbiome [10] [11]. Out of the drugs classes known to alter the composition of the microbiome, most commonly used classes in NSCLC patients were selected. The drug classes selected were antibiotics, proton-pump inhibitors, metformin, antidiabetics and opioids. Medication reconciliation for NSCLC patients was performed by the treating oncologist of assisting nurse prior to the start of the systemic treatment and was reported in the medical health record. Exposure was defined if any information on the use of these drugs was reported within a timeframe of 30 days before and 30 days after the start of either immunotherapy or chemotherapy. This timing was chosen because there is evidence that the effects are the strongest when used shortly before and after the start of immunotherapy and consequently, we would expect the largest effect modification by chemotherapy vs immunotherapy (if any) [17]. The 30 day timeframe is also in line with many previous reports on the topic [9] [13] [18].

### Clinical outcomes

OS was defined as the time from the start of systemic therapy to death. Patients still alive at the end of follow-up on 1 January 2020 were censored at this date.

### Statistical analysis

Statistical Software (SPSS version 26 for Windows: IBM) was used for statistical analysis. Categorical and continuous variables were summarized using descriptive statistics. To compare the immunotherapy and chemotherapy group, we used chi-squared tests (categorical variables) and independent t- test (continuous variables). The potential impact of concomitant medication on OS was analyzed through multivariable cox regression analyses. Possible factors associated with OS were first identified using a univariable analysis. All univariate predictors with a p-value ≤ 0.15 and three other relevant variables - type of treatment, the ECOG-PS and the use of antibiotics - were used to construct the multivariable model. In the final models, backward selection was applied to eliminate non-significant variables (p-value ≤ 0.10). Finally, the models were examined for the existence of effect modification by statistical testing of an interaction term between concomitant drugs of interest and the type of treatment (chemotherapy or immunotherapy). In order to investigate the difference between the lines of treatment, an exploratory analysis was performed for 1L patients and for 2L + 3L patients using the same approach described above. Survival curves using the Kaplan-Meier method were constructed to visualize - where considered of relevance- contrast between the use of concomitant medication and the type of treatment (chemotherapy or immunotherapy).

### Ethical statement

All methods were carried out in accordance with relevant guidelines and regulations. Our ethics committee - the Santeon Institutional Review Board - approved the study (SDB 2019-013) and waived the need for informed consent because of anonymous data handling and the retrospective nature of the study. The study was performed in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

## Results

### Baseline characteristics

A total of 221 immunotherapy patients could be matched to patients treated with chemotherapy in the pre-immunotherapy era resulting in a total of 442 patients available for our primary analysis. The baseline characteristics of the immunotherapy and chemotherapy group are summarized in Table 1. Both immuno- and chemotherapy groups had an average age of 64 years. The majority of patients was male (59.2%) and received immunotherapy as second or higher line of treatment (62%). The demographic and baseline characteristics of the patients were well balanced between the two treatment groups, with the exception of the ECOG-PS. In the immunotherapy group, there was a significantly higher proportion of patients with an ECOG -PS1 (62.5% versus 39.8%) as compared to the chemotherapy group. In both groups, the most common histology subtype was adenocarcinoma, although there were fewer patients in the immunotherapy group with squamous tumor histology than in the chemotherapy group (10.9% versus 18.9% respectively). Table 2 provides a list of concomitant medication in use at the time of start systemic treatment. The most commonly used drugs in both groups were PPIs (43.4% vs 45.7% for the immuno- and chemotherapy group). Patients who received immunotherapy used significantly less opioids (26.7% vs 37.6% p = 0.02) and used more antibiotics as compared to the patients who received chemotherapy (15.8% vs 19.5%, p = 0.32). Only a small percentage of patients in the immunotherapy and chemotherapy group used metformin (3.6% and 9.5% respectively) or other antidiabetics (2.3% and 3.2% respectively).

**Table 1.** Patient characteristics of stage IV NSCLC patients treated with immunotherapy and chemotherapy.

| **Characteristics** | | **Immunotherapy (n=221)** | **Chemotherapy (n=221)** | ***p- value*** |
|---|---|---|---|---|
| Age (median +/- SD) | | 64.0 +/- 9.0 | 64.1 +/- 8.9 | 0.83 |
| Gender | | 131 (59.2%) | 131 (59.2%) | 1.0 |
| BMI (median +/- SD) | | 25.0 +/- 6.3 | 25.0 +/- 4.1 | 0.82 |
| Line of treatment | | | | 1.0 |
| | 1 | 84 (38.0%) | 84 (38.0%) | |
| | 2 | 120 (54.3%) | 120 (54.3%) | |
| | 3 | 17 (7.7%) | 17 (7.7%) | |
| ECOG PS | | | | <0.001 |
| | 0 | 70 (31.7%) | 104 (47.1%) | |
| | 1 | 138 (62.4%) | 88 (39.8%) | |
| | 2 | 10 (4.5%) | 19 (8.6%) | |
| | ≤3 | 2 (0.9%) | 4 (1.8%) | |
| | unknown | 1 (0.5%) | 6 (2.7%) | |
| Immunotherapy | | | | |
| | Nivolumab | 121 (54.8%) | | |
| | Pembrolizumab | 93 (42.1%) | | |
| | Atezolizumab | 7 (3.0%) | | |
| Chemotherapy | | | | |
| | Cisplatin/pemetrexed | | 14 (6.3%) | |
| | Cisplatin/gemcitabine | | 5 (2.3%) | |
| | Carboplatin/pemetrexed | | 57 (31.3%) | |
| | Carboplatin/gemcitabine | | 17 (7.7%) | |
| | Carboplatin/paclitaxel/bevacizumab | | 21 (9.5%) | |
| | Pemetrexed monotherapy | | 46 (20.8%) | |
| | Docetaxel monotherapy | | 58 (26.2%) | |
| | Other | | 3 (1.4%) | |
| Histology | | | | <0.001 |
| | Adenocarcinoma | 176 (79.6%) | 132 (59.7%) | |
| | Squamous | 24 (10.9%) | 40 (18.1%) | |
| | Large cell carcinoma | 10 (4.5%) | 17 (7.7%) | |
| | Other | 11 (4.9%) | 32 (14.5%) | |

**Table 1.** Continued.

| **Characteristics** | | **Immunotherapy (n=221)** | **Chemotherapy (n=221)** | ***p- value*** |
|---|---|---|---|---|
| PD-L1 expression | | | | |
| | <1% | 34 (15.4%) | | |
| | 1-50% | 28 (12.7%) | | |
| | >50% | 100 (45.2%) | | |
| | Unknown | 60 (27.1%) | | |
| Targetable mutations | | | | 0.58 |
| | EGFR | 5 (2.3%) | 6 (2.7%) | |
| | ALK | 0 | 1 (0.5%) | |
| | Other | 5 (2.3%) | 0 | |
| Brain metastasis (yes) | | 49 (22.2%) | 43 (19.5%) | 0.48 |

Abbreviations: ECOG PS, Eastern Cooperative Group performance status; BMI, Body mass index; PD-L1, programmed death ligand-1.

**Table 2.** Concomitant medication exposure according to immunotherapy of chemotherapy treatment

| | **Immunotherapy (n=221)** | **Chemotherapy (n=221)** | ***p-value*** |
|---|---|---|---|
| Antibiotics | 35 (15.8%) | 43 (19.5%) | 0.32 |
| Antidiabetics | 5 (2.3%) | 7 (3.2%) | 0.56 |
| Metformin | 8 (3.6%) | 21 (9.5%) | 0.12 |
| PPI | 96 (43.4%) | 101 (45.7%) | 0.63 |
| Opioids | 61 (27.6%) | 83 (37.6%) | 0.02 |

Abbreviations: PPI, proton pump inhibitor

## Overall survival outcomes in the total population

The results of the univariable and multivariable analysis for OS are summarized in Table 3. The multivariable model showed that the use of antibiotics and opioids was significantly associated with shorter OS, with a corresponding HR of 1.39 (95%CI 1.06 – 1.81) and a HR 1.33 (95%CI 1.07 – 1.66) respectively. Proton pump inhibitors, metformin and other antidiabetics were not associated with OS. Overall, the strongest factor associated with OS was the type of systemic treatment (immunotherapy vs chemotherapy). Other factors identified as independently associated with OS in the multivariable model were type of histology (squamous vs. non-squamous) and line of treatment (≥2L vs. 1L).

For both antibiotics and opioids, the tests for interaction with treatment type were insignificant, both with a p value of 0.50. The distinctive effects of antibiotics and opioids on OS for the immuno- and chemotherapy cohort were summarized in table 4a. The calculated point estimates for the hazard ratio for use of antibiotics towards OS were similar with immunotherapy (HR 1.20 95% CI 0.79 – 1.85) and chemotherapy (HR 1.46 95%CI 1.04 – 2.05). The same was observed for the use of opioids (HR 1.44 (95%CI 1.01 – 2.06) in immunotherapy group and a HR of 1.24 (95%CI 0.94 – 1.63) in chemotherapy group).

**Table 3.** Univariable and multivariable model for overall survival in the total population

| **Overall survival for the total population (n =442)** | **Univariable model** | | | **Multivariable model** | | |
|---|---|---|---|---|---|---|
| | *HR* | *95% CI* | *p-value* | *HR* | *95% CI* | *p-value* |
| Treatment immunotherapy (vs chemotherapy) | 0.45 | 0.36 - 0.56 | <0.01 | 0.46 | 0.37 - 0.58 | <0.01 |
| Histology squamous (vs non-squamous) | 1.38 | 1.04 - 1.84 | 0.03 | 1.33 | 0.10 - 1.79 | 0.05 |
| ECOG PS | | | | | | |
| 1 (vs 0) | 1.00 | 0.80 – 1.25 | 0.99 | 1.08 | 0.86 – 1.35 | 0.52 |
| ≥ 2 (vs 0) | 1.40 | 0.95 – 2.06 | 0.09 | 1.24 | 0.84 – 1.83 | 0.28 |
| Gender female (vs. male) | 0.97 | 0.78 - 1.19 | 0.75 | | | |
| Age ≥75 year ( vs. <75 year) | 0.80 | 0.55 - 1.16 | 0.24 | | | |
| BMI ≥ 25 (vs. <25) | 0.99 | 0.76 - 1.28 | 0.92 | | | |
| Brain metastases yes (vs. no) | 1.08 | 0.84 - 1.39 | 0.53 | | | |
| Line of treatment ≥2 (vs. 1) | 1.45 | 1.17 - 1.80 | <0.01 | 1.44 | 1.16 - 1.78 | <0.01 |
| Antibiotics use (vs. no use) | 1.32 | 1.01 - 1.71 | 0.04 | 1.39 | 1.06 - 1.81 | 0.02 |
| Antidiabetics use (vs. no use) | 1.57 | 0.86 - 2.87 | 0.14 | 1.37 | 0.74 - 2.52 | 0.32 |
| Metformin use (vs. no use) | 1.33 | 0.89 - 1.98 | 0.17 | | | |
| PPI use (vs. no use) | 1.16 | 0.94 - 1.42 | 0.17 | | | |
| Opioid use (vs. no use) | 1.48 | 1.19 - 1.84 | <0.01 | 1.33 | 1.07 - 1.66 | 0.01 |

Abbreviations: ECOG PS, Eastern Cooperative Group performance status; BMI, body mass index PPI, proton pump inhibitor; HR, Hazard ratio; CI, confidence interval

## Overall survival outcomes per treatment line

Restriction of the cohort to patients with 1L treatment yielded similar HRs in that setting. The HRs were 1.49 (95% CI 0.97 – 2.29) for antibiotic use and 1.58 (95% CI 1.08 – 2.32) for opioid use (Supplementary table S1). Also in this setting there was no association between OS and proton pump inhibitors, metformin and other antidiabetics. Factors significantly associated with OS in the multivariable model were the type of treatment (immunotherapy vs chemotherapy) and the presence of brain metastases. The tests for interaction between treatment type and the use of antibiotics and opioids were both not statistically significant, with corresponding p values of 0.89 and 0.15 respectively. The negative effects of antibiotics and opioids on OS were observed in patients treated with chemotherapy as well as in patients treated with immunotherapy (table 4b), with HRs of similar magnitude between the immuno- and chemotherapy group, although not all were statistically significant. In a second line setting, the use of antibiotics and the use opioids were not statistically significantly associated with OS. The multivariable model for OS showed that only the type of treatment (immunotherapy vs chemotherapy) was associated with a lower risk of death (Supplementary table S2). Figures 1a and 1b demonstrate the Kaplan-Meier survival curves for patients treated with immuno- versus chemotherapy, and using/not using antibiotics (1a) or using/not using opioids (1b) in the 1L.

**Table 4.** Multivariable model for OS within the total population (a) and the 1L (b) for both antibiotic and opioid exposure according to type of treatment (chemotherapy or immunotherapy)

| | | |
|---|---|---|
| **a. OS total population** | | |
| **Antibiotic use**[a,1] | | |
| | HR (95% CI) for antibiotic use (vs no use) according to the type of treatment | p- value |
| Chemotherapy cohort | 1.46 (1.04 – 2.05) | 0.03 |
| Immunotherapy cohort | 1.20 (0.79 – 1.85) | 0.39 |
| **Opioid use**[a,2] | | |
| | HR (95% CI) for opioid use (vs no use) according to the type of treatment | p- value |
| Chemotherapy cohort | 1.24 (0.94 – 1.63) | 0.13 |
| Immunotherapy cohort | 1.44 (1.01 – 2.06) | 0.04 |
| **b. OS first line** | | |
| **Antibiotic use**[b.3.] | | |
| | HR (95% CI) for antibiotic use (vs no use) according to the type of treatment | p- value |
| Chemotherapy cohort | 1.54 (0.86 – 2.76) | 0.15 |
| Immunotherapy cohort | 1.45 (0.75 – 2.77) | 0.27 |
| **Opioid use**[b.4.] | | |
| | HR (95% CI) for opioid use (vs no use) according to the type of treatment | p- value |
| Chemotherapy cohort | 1.30 (0.80 – 2.09) | 0.29 |
| Immunotherapy cohort | 2.27 (1.24 – 4.16) | < 0.01 |

The multivariable model includes the variables:

[a] *Treatment (immunotherapy vs chemotherapy), histology (squamous vs non-squamous), line of treatment (≥2 vs. 1), antibiotic (use vs no use), opioids (use vs no use).*

[1.] *Interaction between antibiotic use and type of treatment*

[2.] *Interaction between opioid use and type of treatment*

[b.] *Treatment (immunotherapy vs chemotherapy), brain metastases (yes vs no), antibiotic (use vs no use), opioids (use vs no use).*

[3] *Interaction between antibiotic use and type of treatment*

[4] *Interaction between opioid use and type of treatment*

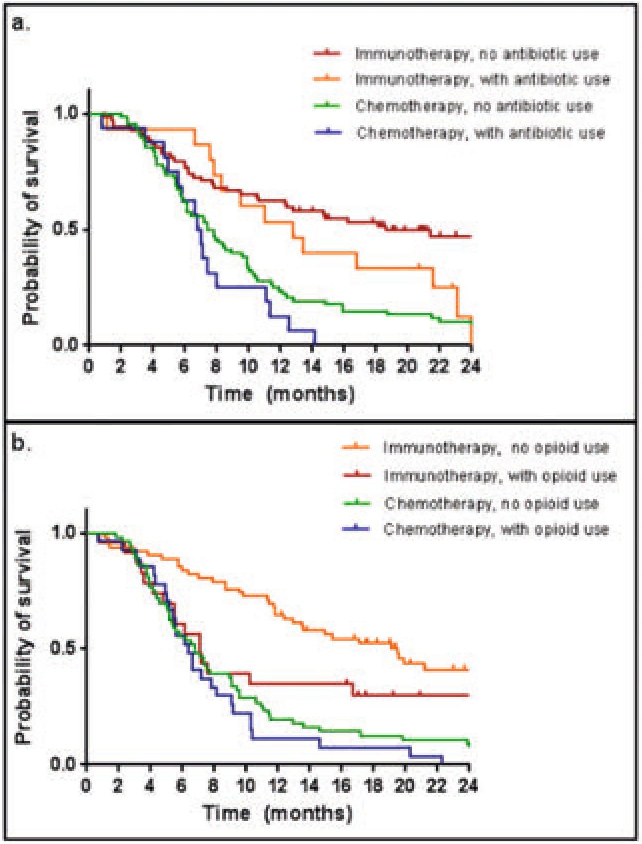


**Figure 1.** Kaplan-Meier curves showing overall survival for first-line chemo- and immunotherapy with and without the use of antibiotics (a) and with and without the use of opioids (b).

## Discussion

In this historically matched cohort study of 442 stage IV NSCLC patients, the use of antibiotics and opioids were shown to be independently associated with worse survival outcomes after adjusting for other prognostic factors associated with a poorer prognosis. Moreover, the negative influence of antibiotics and opioids on OS was observed both for treatment with chemotherapy as well as with immunotherapy. This finding suggests that the association between worse survival outcomes and the use of antibiotics and opioids more likely originates from its use being linked with confounding factors rather than the disturbing effect these drugs have on the gut microbiome.

We believe that our historically matched cohort study design adds to what has been published about the topic so far because all previously published reports on the association of antibiotics with immunotherapy effectiveness are retrospective single arm cohort studies. Lurienne et al (2020) summarized these cohort studies (n=21) in a meta-analysis and reported a pooled HR for OS of 1.69 (95% CI 1.25- 2.29) for NSCLC patients exposed to antibiotics when starting immunotherapy [17]. Our observed HR of 1.39 is in line with this pooled number but appears of similar magnitude in patients treated with chemotherapy in the years that immunotherapy was not yet available. This finding argues against causality because an effect of antibiotics on chemotherapy effectiveness as a result of microbiome disturbance is unlikely. Thus, we consider that residual confounding by indication bias plays a role in previous reports. An exception is the recent study of Chalabi et al. In a post hoc analysis of a randomized clinical trial of atezolizumab versus docetaxel in a second line setting, a larger association between antibiotics and worse OS was observed in the atezolizumab study arm compared to chemotherapy [18]. However, also in that study residual confounding cannot be excluded. It is also conceivable that symptomatic rapid progressive disease represented by antibiotic use, prevented patients randomized to immunotherapy from early disease control resulting in earlier death eventually [19]. Finally, Cortellini et al recently reported a non-significant HR of 1.42 for antibiotics in patients treated with the combination of immunotherapy and chemotherapy in first line. This HR aligns very much with our observed HR in patients receiving chemotherapy (1.54), further supporting the hypothesis that recent antibiotic use is primarily a prognostic factor instead of an etiological factor [20].

For opioid use our findings align with what has been studied by Zheng et al 2020 [21]. The authors conducted a cohort analysis (n =203 patients) and a meta-analysis (n = 26 articles) to investigate the impact of opioids use on survival of cancer patients. The results of their analyses showed a negative association between cancer-specific survival and the use of opioids explained cancer-related pain treated with opoids. Furthermore, other studies have reported that opioids may negatively affect cancer patients' survival through respiratory depression, delirium, addiction or directly by acting on tumor cells [22] [23]. Even though these authors did not investigate whether the negative effects of opioids on survival outcomes differ between patients treated with immunotherapy or chemotherapy, it supports our observed association between opioids and survival irrespective of the systemic treatment applied.

Noteworthy is that the use of PPIs, metformin and other antidiabetics were not associated with poor survival outcomes in our cohort. This might indicate that both the microbiome-hypothesis and the hypothesis of confounding by indication should be rejected for these drugs. Interesting, however, is that some other studies observed associations between PPI use and worse outcomes [16]. That we were unable to replicate this might be explained by the very high prevalence of PPI use in our cohort that could have resulted in overshadowing an assumed proxy of PPI use for worse clinical condition. In the Netherlands, PPIs are recommended for every adult >65 years and using anti-platelet drugs, NSAIDS or corticosteroids. Nevertheless, our data suggest no effect of PPIs on effectiveness of immunotherapy in NSCLC patients. The frequency of patients using metformin (6.5%) and antidiabetics (2.7%) in our cohort we consider too small to conclude on any association between these drugs and survival outcomes.

## Strengths and weaknesses

Our study has a number of notable strengths. The main strength is its matched cohort study design. This design enabled us to investigate the possibility of concomitant medication-use having a prognostic impact on patients' clinical survival, regardless of the treatment given, under the assumption that unmeasured confounding factors will be balanced between the two groups. That we captured chemotherapy patients from a period that immunotherapy was not yet available maximized the potential for balance because after market access of immunotherapy patient characteristics are more likely to guide the decision for either treatment. Nevertheless, the distribution of the ECOG status appeared imbalanced in our study. The most likely explanation for this is that the milder toxicology from immunotherapy has led to patients who would not qualify for toxic chemotherapy starting with immunotherapy anyway. By fixing the ECOG status in all our multivariable regression models we think that adjustment for this imbalance has been maximized. Another positive aspect is that we were able to evaluate not only the effects of antibiotics on the therapeutic outcomes of NSCLC patients but also the effects of other drugs, known to be associated with changes of the microbiome composition. There are numerous observational studies investigating the impact of antibiotics, however, studies regarding the effects of other drugs are scarce. Lastly, compared to previous reports, we used a relatively large cohort with advanced NSCLC patients treated in six different hospitals across the Netherlands.

Limitation of our study is the small number of patients using some of the drugs under study. This might explain some null findings due to limited statistical power. Another limitation, in hindsight, is that we did not collect data about the use of corticosteroids in our patients. Corticosteroids have been linked to altered immunotherapy effectiveness as well and inclusion of this variable might have resulted in other HRs with our variables of interest [24] [25]. Some co-linearity of corticosteroids and antibiotics or PPIs cannot be ruled out. Another potential drawback is that we examined concomitant medication when used within a relative short time frame of 30 days before up to 30 days after the start of the systemic therapy. Although the decision for this time frame was well-thought-out (see Methods), we acknowledge that this focus prevented us from studying time varying associations. Finally, our study shares all limitations linked to retrospective studies, at least information bias. For example, the use of concomitant medication was fully based on the information recorded in the patients' health records and may not be completely comprehensive. On the other hand, both patients with chemotherapy and with immunotherapy were captured from the same hospitals, making the risk of misclassification bias, if any, being non-differential.

## Conclusion

In conclusion, we observed that the use of antibiotics and opioids is associated with shorter survival, similarly in chemotherapy and immunotherapy-treated NSCLC patients. This suggests that the association is likely to be a consequence of confounding rather than disturbing the composition of the microbiome.

## References

1. Union for International Cancer Control. Non-small cell lung cancer - 2014 Review of Cancer Medicines on the WHO List of Essential Medicines. https://www.who.int/selection_medicines/committees/expert/20/applications/NonSmallCellLungCancer.pdf?ua=1 (2014)
2. Chen R, Manochakian R, James L, Azzouqa AG, Shi H, Zhang Y, et al. Emerging therapeutic agents for advanced non-small cell lung cancer. J Hematol Oncol. 2020;13(1):1–23.
3. Kastelijn EA, de Langen AJ, Peters BJM. Treatment of oncogene-driven non-small cell lung cancer. Curr Opin Pulm Med. 2019;25(3):300–7.
4. Brahmer J, Reckamp KL, Baas P, Crinò L, Eberhardt WEE, Poddubskaya E, et al. Nivolumab versus docetaxel in advanced squamous-cell non–small-cell lung cancer. N Engl J Med. 2015;373(2):123–35.
5. Rittmeyer A, Barlesi F, Waterkamp D, Park K, Ciardiello F, von Pawel J, et al. Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (OAK): a phase 3, open-label, multicentre randomised controlled trial. Lancet. 2017;389(10066):255–65.
6. Herbst RS, Baas P, Kim DW, Felip E, Pérez-Gracia JL, Han JY, et al. Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): A randomised controlled trial. Lancet. 2016;387(10027):1540–50.
7. Iida N, Dzutsev A, Stewart CA, Smith L, Bouladoux N, Weingarten RA, et al. Commensal bacteria control cancer response to therapy by modulating the tumor microenvironment. Science. 2013;342(6161):967–70.
8. Hooper L V., Littman DR, Macpherson AJ. Interactions between the microbiota and the immune system. Science. 2012;336(6086):1268–73.
9. Routy B, Le Chatelier E, Derosa L, Duong CPM, Alou MT, Daillère R, et al. Gut microbiome influences efficacy of PD-1-based immunotherapy against epithelial tumors. Science. 2018;359(6371):91–7.
10. Maier L, Pruteanu M, Kuhn M, Zeller G, Telzerow A, Anderson EE, et al. Extensive impact of non-antibiotic drugs on human gut bacteria. Nature. 2018;555(7698):623–8. Available from: http://dx.doi.org/10.1038/nature25979
11. Vich Vila A, Collij V, Sanna S, Sinha T, Imhann F, Bourgonje AR, et al. Impact of commonly used drugs on the composition and metabolic function of the gut microbiota. Nat Commun. 2020;11(1):1–11.
12. Zhao S, Gao G, Li W, Li X, Zhao C, Jiang T, et al. Antibiotics are associated with attenuated efficacy of anti-PD-1/PD-L1 therapies in Chinese patients with advanced non-small cell lung cancer. Lung Cancer. 2019;130:10–7.
13. Derosa L, Hellmann MD, Spaziano M, Halpenny D, Fidelle M, Rizvi H, et al. Negative association of antibiotics on clinical activity of immune checkpoint inhibitors in patients with advanced renal cell and non-small-cell lung cancer. Ann Oncol. 2018;29(6):1437–44.
14. Van De Garde E, Plouvier B, Fleuren H, et al. Pharmacotherapy within a learning healthcare system: rationale for the Dutch Santeon Farmadatabase. Eur J Hosp Pharm. 2019;26(1):46-50
15. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)-a metadata-driven methodology and workflow process for providingtranslational research informatics support. J Biomed Inform. 2009;42(2):377-81.
16. Cramer-Van Der Welle CM, Peters BJM, Schramel FMNH, Klungel OH, Groen HJM, Van De Garde EMW. Systematic evaluation of the efficacy-effectiveness gap of systemic treatments in metastatic nonsmall cell lung cancer. Eur Respir J. 2018;52(6).
17. Lurienne L, Cervesi J, Duhalde L, de Gunzburg J, Andremont A, Zalcman G, et al. NSCLC immunotherapy efficacy and antibiotic use: A systematic review and meta-analysis. J Thorac Oncol. 2020;15(7):1147–59.

18. Chalabi M, Cardona A, Nagarkar DR, Dhawahir Scala A, Gandara DR, Rittmeyer A, et al. Efficacy of chemotherapy and atezolizumab in patients with non-small-cell lung cancer receiving antibiotics and proton pump inhibitors: pooled post hoc analyses of the OAK and POPLAR trials. Ann Oncol. 2020;31(4):525–31.
19. van de Garde EMW, van Bedaf LR, Hurkmans DP, van den Heuvel MM. Antibiotic use and reduced effectiveness of second-line immunotherapy for lung cancer: all the time or just at the start of treatment? Ann Oncol. 2020;31(12):1779–80.
20. Cortellini A, Ricciuti, B, Facchinetti F, et al. Antibiotic-exposed patients with non-small-cell lung cancer preserve efficacy outcomes following first-line chemo-immunotherapy. Ann Oncol [Preprint] 2021. Available from: 10.1016/j.annonc.2021.08.1744 10.
21. Zheng J, He J, Wang W, Zhou H, Cai S, Zhu L, et al. The impact of pain and opioids use on survival in cancer patients: Results from a population-based cohort study and a meta-analysis. Med (United States). 2020;99(9).
22. Hasegawa T, Oguri T, Osawa T, Sawa T, Osaga S, Okuyama T, et al. Opioid dose and survival of patients with incurable nonsmall cell lung cancer: A prospective cohort study. J Palliat Med. 2018;21(10):1436–41.
23. Singleton PA, Moss J, Karp DD, Atkins JT, Janku F. The mu opioid receptor: A new target for cancer therapy? Cancer. 2015;121(16):2681–8.
24. Drakaki A, Dhillon PK, Wakelee H, Chui SY, Shim J, Kent M, et al. Association of baseline systemic corticosteroid use with overall survival and time to next treatment in patients receiving immune checkpoint inhibitor therapy in real-world US oncology practice for advanced non-small cell lung cancer, melanoma, or urothel. Oncoimmunology. 2020;9(1):1–9.
25. Mountzios G, de Toma A, Economopoulou P, Friedlaender A, Banini M, Lo Russo G, et al. Steroid use independently predicts for poor outcomes in patients with advanced NSCLC and high PD-L1 expression receiving first-line pembrolizumab monotherapy. Clin Lung Cancer. 2020.

# Supplementary

**Table S1.** Univariable and multivariable model for overall survival in first-line treatment

| **Overall survival for first-line treatment (n = 168)** | | | | | | |
|---|---|---|---|---|---|---|
| | **Univariable model** | | | **Multivariable model** | | |
| | **HR** | **95% CI** | **p-value** | **HR** | **95% CI** | **p-value** |
| Treatment immunotherapy (vs chemotherapy) | 0.37 | 0.26 - 0.54 | <0.01 | 0.37 | 0.26 - 0.54 | <0.001 |
| Histology squamous (vs non-squamous) | 1.29 | 0.80 - 2.08 | 0.30 | | | |
| ECOG PS | | | | | | |
| 1 (vs 0) | 1.06 | 0.73 – 1.53 | 0.76 | 1.02 | 0.70 – 1.48 | 0.92 |
| ≥ 2 (vs 0) | 1.59 | 0.86 – 2.93 | 0.17 | 1.55 | 0.83 – 2.90 | 0.17 |
| Gender female (vs. male) | 1.10 | 0.78 - 1.56 | 0.60 | | | |
| Age ≥75 year ( vs. <75 year) | 0.85 | 0.51 - 1.41 | 0.85 | | | |
| BMI ≥ 25 (vs. <25) | 1.09 | 0.72 - 1.65 | 0.68 | | | |
| Brain metastases yes (vs. no) | 1.46 | 0.99 - 2.14 | 0.06 | 1.54 | 1.04 - 2.28 | 0.03 |
| Antibiotics use (vs. no use) | 1.46 | 0.96 - 2.21 | 0.08 | 1.49 | 0.97 - 2.29 | 0.07 |
| Antidiabetics use (vs. no use) | 2.14 | 0.53 - 8.70 | 0.29 | | | |
| Metformin use (vs. no use) | 1.77 | 0.10 - 3.15 | 0.05 | 1.07 | 0.58 - 1.97 | 0.83 |
| PPI use (vs. no use) | 1.24 | 0.87 - 1.75 | 0.24 | | | |
| Opioids use (vs. no use) | 1.75 | 1.20 - 2.54 | <0.01 | 1.58 | 1.08 - 2.32 | 0.02 |

Abbreviations: ECOG PS, Eastern Cooperative Group performance status; BMI, body mass index PPI, proton pump inhibitor; pump inhibitor; HR, Hazard ratio; CI, confidence interval.

**Table S2.** Univariable and multivariable model for overall survival in second- and third-line treatment

| **Overall survival for second- and third line treatment (n = 274)** | | | | | | |
|---|---|---|---|---|---|---|
| | **Univariable model** | | | **Multivariable model** | | |
| | **HR** | **95% CI** | **p-value** | **HR** | **95% CI** | **p-value** |
| Treatment immunotherapy (vs chemotherapy) | 0.51 | 0.39 - 0.66 | <0.01 | 0.52 | 0.40 - 0.68 | <0.01 |
| Histology squamous (vs non-squamous) | 1.47 | 1.03 - 2.10 | 0.03 | 1.34 | 0.93 - 1.94 | 0.12 |
| ECOG PS | | | | | | |
| 1 (vs 0) | 0.98 | 0.74 – 1.29 | 0.87 | 1.10 | 0.82 – 1.48 | 0.51 |
| ≥2 (vs 0) | 1.31 | 0.80 – 2.15 | 0.29 | 1.11 | 0.66 – 1.86 | 0.69 |
| Gender female (vs. male) | 0.91 | 0.70 - 1.18 | 0.47 | | | |
| Age ≥75 year ( vs. <75 year) | 0.89 | 0.51 - 1.56 | 0.68 | | | |
| BMI ≥ 25 (vs. <25) | 0.85 | 0.60 - 1.19 | 0.34 | | | |
| Brain metastases yes (vs. no) | 0.89 | 0.64 - 1.25 | 0.52 | | | |
| Antibiotics use (vs. no use) | 1.28 | 0.91 - 1.80 | 0.16 | 1.30 | 0.92 - 1.84 | 0.14 |
| Antidiabetics use (vs. no use) | 1.29 | 0.66 - 2.52 | 0.46 | | | |
| Metformin use (vs. no use) | 1.11 | 0.63 - 1.94 | 0.73 | | | |
| PPI use (vs. no use) | 1.04 | 0.81 - 1.35 | 0.75 | | | |
| Opioids use (vs. no use) | 1.31 | 1.00 - 1.71 | 0.05 | 1.23 | 0.94 - 1.61 | 0.14 |

Abbreviations: ECOG PS, Eastern Cooperative Group performance status; BMI, body mass index PPI, proton pump inhibitor; pump inhibitor; HR, Hazard ratio; CI, confidence interval

# CHAPTER 3

# Methodological challenges and advancements for prompt evidence generation

# 3.1

# Definition, measurement and reporting of progression-free survival in randomized clinical trials and observational studies of patients with advanced non-small cell lung cancer treated with immunotherapy: a scoping review.

Marjon V Verschueren, Vasiliki Tassopoulou, Robbert Visscher,
Jeroen Schuurkamp, Bas JM Peters, Miriam Koopman,
Ewoudt MW van de Garde, Antoine CG Egberts, Lourens T Bloem.

*ESMO real-world data and digital oncology.2025;7:100118.*
*DOI: 10.1016/j.esmorw.2025.100118*

# Abstract

## Background

Evidence from observational studies is increasingly used in oncology to complement evidence from clinical trials. Commonly used endpoints to evaluate oncology medicines are overall survival (OS) and progression-free survival (PFS). However, comparing PFS across observational studies and with clinical trials can be challenging due to differences in its definition and measurement. This scoping review investigated how PFS was defined, measured and reported in randomized clinical trials (RCTs) and observational studies of patients with advanced non-small cell lung cancer (NSCLC) treated with immunotherapy.

## Methods

This scoping review included RCTs and observational studies that measured PFS in advanced NSCLC patients treated with immunotherapy. ASReview, an open-source artificial intelligence-assisted tool, was used to screen and prioritize relevant studies from records identified from PubMed and Embase between 2012-2023. Information on study characteristics, PFS definitions and measurements were extracted.

## Results

40 RCTs and 144 observational studies were included. Most RCTs were conducted across multiple continents (70%), while most observational studies were conducted in Asia (62%). In contrast to RCTs, many observational studies lacked reporting on the end date of PFS measurement (69%), the type of radiological imaging (59%), and the imaging reviewer (78%). For observational studies that did report on PFS definitions and measurements, these often differed from those in RCTs, particularly regarding event definitions, the start- and stop dates for PFS measurement, and tumor assessment schedules.

## Conclusions

In contrast to RCTs, observational studies often lack reporting on PFS definitions and measurements, and if reported, they differ across observational studies and between them and RCTs. Since observational studies are important for complementing evidence, aligning PFS definition and measurement criteria with those used in RCTs, along with detailed reporting, is needed. However, some variability in PFS measurement characteristics is unavoidable, and therefore, PFS estimates from observational studies should be interpreted critically and carefully.

## Introduction

Evidence from observational studies is increasingly used to complement evidence from clinical trials to understand the benefits and risks of new oncology medicines because the generalizability of these trials to daily clinical practice is often limited [1]. Patients in randomized clinical trials (RCTs) typically constitute a highly selective sample, are often younger, fitter and with fewer comorbidities than the broader patient population treated in daily clinical practice. In addition, the variability in treatment adherence and outcome assessments is much greater in clinical practice compared to the strictly defined and controlled conditions of RCTs [2,3].

An important aspect when using evidence from observational studies to complement that from RCTs, is the definition, measurement, and reporting of endpoints. Over the past decade, the primary outcome in evaluating oncology medicines has shifted from overall survival (OS) to progression-free survival (PFS), since PFS is not influenced by subsequent treatment and requires shorter follow-up periods than OS [4]. However, PFS is subject to inconsistencies in its definition and measurement, among others, due to the use of varying criteria (e.g., Response Evaluation Criteria in Solid Tumors, RECIST, immunotherapy-specific iRECIST or clinical criteria) and varying response assessment intervals [5-8]. These inconsistencies may lead to estimates of PFS that are not comparable, either within or between studies. For example, if one treatment arm is assessed more frequently than another within the same study, disease progression might be detected earlier, leading to biased absolute and relative PFS estimates.This issue extends to comparisons between studies, especially between observational studies and RCTs, with their often inherent differences in data collection.

These issues underscore the need for standardized reporting for observational studies to enhance the interpretability and reliability of PFS outcomes. In addition to existing general guidelines on reporting standards for observational studies [9,10], the European Society for Medical Oncology (ESMO) and the American Society of Clinical Oncology (ASCO) have recently introduced specific guidance to oncology research [11,13]. Despite these developments, the degree of variation in how PFS is defined, measured, and reported across different study types remains unclear. Therefore, this scoping review aimed to evaluate how PFS was defined, measured, and reported for RCTs and observational studies of patients with unresectable advanced non-small cell lung cancer (NSCLC) treated with immunotherapy. The population of unresectable advanced NSCLC was chosen due to its relatively short PFS and the use of immunotherapy as a standard treatment option.

## Methods

This scoping review was conducted in accordance with the Joanna Briggs Institute manual for scoping reviews [13] and the checklist of Preferred Reporting Items for Systematic Reviews and Meta-analyses – Extension for Scoping Review' (PRISMA-ScR) [14].

### Eligibility criteria

This scoping review aimed to include all scientific publications of phase III RCTs and observational studies of patients with advanced unresectable NSCLC that compared the PFS of at least one PD-1/PD-L1 or CTLA-4 inhibiting immunotherapy treatment with one or more other treatments. For observational studies, this also included comparisons with data from clinical trials, either with trial estimates or with individual patient data. Phase I or II trials, interim analyses, updates, subgroup analyses, or publications not written in English were excluded. .

### Information sources, search strategy and study selection

A systematic search for available literature was carried out in the following two databases: PubMed and Embase. The databases were searched for articles published between January 2012 and March 2023. The full search term for each database is provided in Table S1.

The retrieved study records were first transferred to Rayyan QCRI and deduplicated. Then, the remaining records were transferred to an artificial intelligence-assisted review tool named ASReview. This open-source software systematically screens records with researcher-trained and continuously updated machine-learning algorithms [15]. For this study, we used naïve Bayes for classification and term frequency-inverse document frequency for feature extraction. Two reviewers (MV and RV) independently reviewed the titles and abstracts prioritized by ASReview to identify relevant reports. The reviewers stopped the title and abstract screening when the stopping rule of minimum of 2250 records and 25 subsequent irrelevant records was reached [16]. For both reviewers, the stopping rule was met after screening 2250 records, and the remaining records were labeled irrelevant by the review tool without interference from the reviewers Both authors independently read the full text of the selected reports. Disagreements about inclusion were resolved by discussion with a third reviewer (LB).

### Data extraction

Two reviewers (MV and VT) independently extracted data from the included studies using the data extraction form provided in the Supplementary material (Table S2). Disagreements were resolved by discussion until consensus was reached, or with an additional reviewer (LB).

The following characteristics were extracted: main author, year of publication, publication journal, initiated by pharmaceutical industry, continent, site type (multicenter or single center), type of data, type of comparison (number of treatment-arms for RCTs and number of contemporary or historical cohorts for observational studies), follow-up time for the total study population, stage of disease, histology, line of treatment, type of treatment of interest (any type of PD-1/PD-L1 or CTLA-4 inhibitor alone, or in combination with chemotherapy, targeted agents or other immunotherapy), cycle duration for the treatment of interest, follow-up time for the treatment of interest, comparator treatment (any type of systemic treatment or trial estimate or individual patient data trial), cycle duration for the comparator treatment, follow-up time for the comparator treatment. The following characteristics of PFS measurement were extracted: progression-defining events, start and end dates used for PFS calculation, radiological and/or non-radiological criteria (and if radiological, the type of criteria), type of radiological imaging, radiological imaging reviewer, and tumor assessment schedule. The type of radiological criteria is further categorized based on conventional (RECIST v1.1 introduced in 2009 [17], the newer response criteria (immune related- [18], immune- [19], and immune modified- RECIST [20], respectively introduced in 2013, 2017 and 2018) and a category of other or modified criteria. The characteristic 'Response assessment schedule' was categorized into highly detailed, moderately detailed, and not detailed, as described in Table S3.

### Data synthesis

Descriptive statistics were used to summarize and compare the findings between RCTs and observational studies. The completeness of PFS definitions and measurements reported for each observational study was assessed using a scoring system with one point for each characteristic of PFS measurement, leading to a maximum total score of eight points. For each observational study characteristic, the distribution of reporting was presented in a boxplot.

## Results

### Search strategy and study selection

The initial search yielded a total of 9162 records after deduplication. Of these, 8474 were excluded based on title and abstract screening using the AI-assisted review tool ASReview. Once the stopping rule was met, 688 reports were sought for retrieval; however, 418 could not be retrieved, mainly due to their status as conference abstract from Embase. After assessing 270 full-text reports for eligibility, 86 more reports were excluded. A total of 184 studies were thus included for data extraction. Of these, 40 were RCTs [31-70] and 144 were observational studies [71-214] (Figure 1).

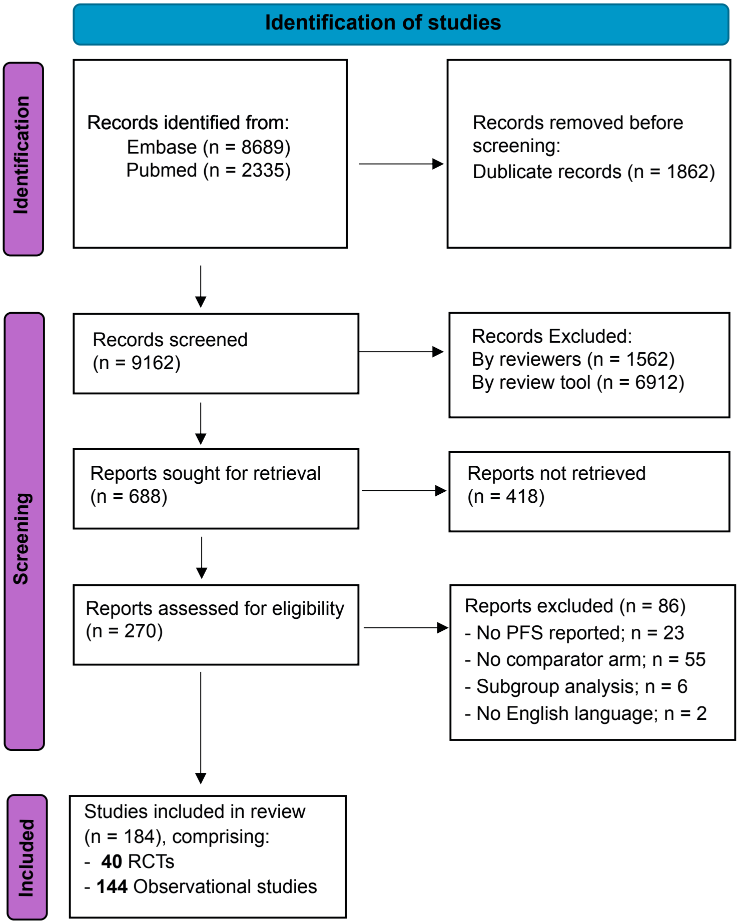


**Figure 1.** Prisma-flow diagram

## Study characteristics

Most included RCTs were published between 2021 and 2023 (48%) in journals with a median impact factor of 45.9 (Interquartile range [IQR] 20.4 – 70.7), Table 1. All RCTs were initiated by the pharmaceutical industry, and were mostly conducted across multiple continents (70%). Most RCTs included patient populations with only stage IV disease (50%), with both non-squamous and squamous histologies (58%), and receiving first-line treatment (80%). In addition, most RCTs compared two treatment arms (90%), with immunotherapy most often as the treatment of interest (38%) and chemotherapy most often as the comparator (62%).

**Table 1.** Characteristics of the included clinical trials and observational studies.

| Characteristics | | RCTs | Observational studies |
|---|---|---|---|
| | | N = 40 | N = 144 |
| **Study information** | | | |
| **Year of publication, n (%)** | | | |
| 2015-2017 | | 7 (18) | 1 (1) |
| 2018-2020 | | 14 (35) | 43 (30) |
| 2021-2023 | | 19 (48) | 100 (69) |
| **Journal impact factor** | | | |
| Median [IQR] | | 45.9 [20.4 – 70.7] | 4.4 [2.9 – 5.3] |
| Not reported | | 0 | 1 (1) |
| **Study design** | | | |
| **Initiated by pharmaceutical industry, n (%)** | | | |
| Yes | | 40 (100) | 5 (3) |
| No | | 0 (0) | 139 (97) |
| **Continent, n (%)** | | | |
| Asia | | 12 (30) | 90 (62) |
| America | | 0 (0) | 36 (25) |
| Europe | | 0 (0) | 12 (8) |
| Australia | | 0 (0) | 3 (3) |
| Multicontinent | | 28 (70) | 1 (1) |
| Not reported | | 0 (0) | 2 (1) |
| **Site type, n (%)** | | | |
| Multicenter | | 40 (100) | 53 (37) |
| Single center | | 0 (0) | 89 (62) |
| Not reported | | 0 (0) | 2 (1) |
| **Type of data, n (%)** | | | |
| Electronic health records | | 0 (0) | 74 (51) |
| EHR combined with other data | | 0 (0) | 13 (9) |
| Registry/database | | 0 (0) | 11 (8)[a] |
| Clinical trial | | 40 (100) | 0 (0) |
| Other | | 0 (0) | 2 (1)[b] |
| Not reported | | 0 (0) | 44 (31) |
| **Type of comparison, n (%)** | | | |
| For trials | Two treatment arms | 36 (90) | NA |
| | >Two treatment arms | 4 (10) | NA |
| For observational studies | One cohort versus trial estimate (descriptive) | NA | 34 (24) |
| | One cohort versus IPD trial | NA | 4 (3) |
| | Two cohorts: Contemporary | NA | 82 (57) |

**Table 1.** Continued.

| **Characteristics** | **RCTs** | **Observational studies** |
|---|---|---|
| | **N = 40** | **N = 144** |
| Two cohorts: Historical | NA | 8 (6) |
| >Two cohort: Contemporary | NA | 16 (11) |
| **Follow-up time in months** | | |
| Total study population (median [IQR]) | 11.7 [9.0 – 16.3] | 13.9 [9.3 – 20.2] |
| Not reported, n (%) | 20 (45) | 63 (44) |
| **Stage of disease, n (%)** | | |
| Stage III[$] | 2 (5) | 15 (10) |
| Stage IV | 20 (50) | 72 (52) |
| Stage III[$] and IV | 18 (45) | 40 (28) |
| Not reported | 0 (0) | 17 (12) |
| **Histology, n (%)** | | |
| Non-squamous | 7 (18) | 47 (33) |
| Squamous | 10 (7) | 2 (1) |
| Both | 23 (58) | 91 (63) |
| Not reported | 0 (0) | 4 (3) |
| **Line of treatment, n (%)** | | |
| 1L | 32 (80) | 36 (25) |
| 2L or later lines | 8 (20) | 52 (36) |
| Different lines | 0 (0) | 42 (29) |
| Not reported | 0 (0) | 14 (10) |
| **Treatment of interest** | | |
| **Type of treatment, n (% is n/N*)** | | |
| Immunotherapy (monotherapy) | 17 (38) | 48 (30) |
| Immunotherapy + chemotherapy | 15 (33) | 42 (26) |
| Immunotherapy + immunotherapy (+ chemotherapy) | 9 (20) | 0 (0) |
| Immunotherapy + (chemo)radiotherapy | 1(2) | 22 (14) |
| Immunotherapy + anti-angiogenesis (+ chemotherapy) | 3 (7) | 22 (14) |
| Other | 0 (0) | 1 (1)[c] |
| Two or more of the above treatments within one arm | 0 (0) | 26 (16) |
| **Cycle duration, n (% = n/N*)** | | |
| Two weeks | 6 (13) | 23 (14) |
| Three weeks | 29 (64) | 28 (18) |
| Other | 9 (20) | 24 (15) |
| Not reported | 0 (0) | 85 (53) |

**Table 1.** Continued.

| **Characteristics** | **RCTs** | **Observational studies** |
|---|---|---|
| | **N = 40** | **N = 144** |
| **Follow-up time in months** | | |
| Median [IQR] | 14.7 [12.2 – 18.9] | 14.3 [9.2 – 18.6] |
| Not reported, n (% is n/N*) | 20 (45) | 135 (84) |
| **Comparator treatment** | | |
| **Type of treatment, n (% = n/N*)** | | |
| Immunotherapy (monotherapy) | 2 (5) | 48 (32) |
| Immunotherapy + chemotherapy | 0 (0) | 3 (2) |
| Anti-angiogenesis (+chemotherapy or immunotherapy) | 1 (3) | 15 (10) |
| Chemotherapy | 25 (62) | 23 (15) |
| (chemo)radiotherapy (+ immunotherapy) | 0 (0) | 14 (9) |
| Placebo (+ combinations) | 12 (30) | 0 (0) |
| Two or more of the above treatments within one arm | 0 (0) | 9 (6) |
| NA (comparison with trial) | NA | 38 (25) |
| **Cycle duration, n (% = n/N*)** | | |
| Two weeks | 2 (5) | 3 (2) |
| Three weeks | 32 (80) | 29 (19) |
| Other | 5 (15) | 12 (8) |
| Not reported | 0 (0) | 106 (71) |
| **Follow-up time in months** | | |
| Median [IQR] | 14.6 [10.9 – 17.8] | 16.5 [12.8 – 29.5] |
| Not reported, n (% is n/N[#]) | 32 (71) | 127 (85) |

Abbreviations: 1L, first-line treatment; 2L, second-line treatment; IQR, interquartile range; IPD, individual patient data; NA, not applicable.

* The total number of treatment arms defined as treatment of interest was n=45 for clinical trials and n=161 for observational studies.

[#] The total number of treatment arms defined as comparator treatment this percentage was n=40 for clinical trials and n=150 for observational studies

[$] Only studies in patients with unresectable stage III disease were included.

[a] One study used interviews and the other study used data reported by physicians.

[b] Seven studies used disease specific registries and four studies did not specify the type of registry used.

[c] Immunotherapy + anti-osteoporosis.

Most observational studies were published between 2021 and 2023 (69%) in journals with a median impact factor of 4.4 [IQR 2.9 – 5.3]. Most observational studies were not initiated by the pharmaceutical industry (97%), and were conducted in Asia (62%), in single centers (62%), and used electronic health records (51%). Most observational studies included patient populations with stage IV disease (54%), with both non-squamous and squamous histologies (63%) and receiving second-line or higher-line treatment (36%). In addition, most observational studies compared treatments between two contemporary cohorts (57%), with immunotherapy most often as the treatment of interest (30%) and as the comparator treatment (32%).

The treatments in the RCTs with two treatment arms and the observational studies with two cohorts are presented in a circle plot, respectively Figure S1 and S2. RCTs most often compared immunotherapy monotherapy with chemotherapy, while observational studies most often compared immunochemotherapy with immunotherapy monotherapy.

## PFS measurement characteristics

All PFS measurement characteristics are presented in Figure 2. For most RCTs, PFS was defined with events including both progression and death (98%), and follow-up from the date of randomization (96%) to the date of the last radiological scan (90%). For measurement criteria, radiological criteria solely were used for most RCTs (98%), with RECIST v.1.1 criteria used in 98% of studies. Disease progression was assessed using (PET-)CT and/or MRI for all RCTs and radiological images were reviewed by a radiologist for 75% of RCTs. The response assessment schedule was highly detailed for all RCTs. In addition, reporting rate for RCTs was high, with fewer than 5% of measurement characteristics not reported.

Reporting for observational studies was low, with significant proportions of studies not providing details on key PFS measurement characteristics: event definition (22%), start date (16%), end date (69%), type of radiological imaging (59%), imaging reviewer (78%), and tumor assessment schedule (57%). When reported, 66% of observational studies defined PFS based on both progression and death events. For the follow-up period, 76% of studies used the date of treatment initiation, and 17% used the date of last follow-up. Regarding measurement criteria, only radiological criteria were used in 73% of the studies, only non-radiological criteria in 2%, and a combination of both in 13%. Among studies using radiological criteria, 70% employed RECIST v1.1. Disease progression was measured using (PET-)CT and/or MRI for 75% of observational studies and radiological images were reviewed by different reviewers. The level of detail for the tumor assessment schedule varied from highly detailed (25%), moderately detailed (13%) and not detailed (6%).

For all observational studies, the completeness of reporting on PFS definitions and measurement characteristics did not vary according to study information and study design characteristics (Table S4).

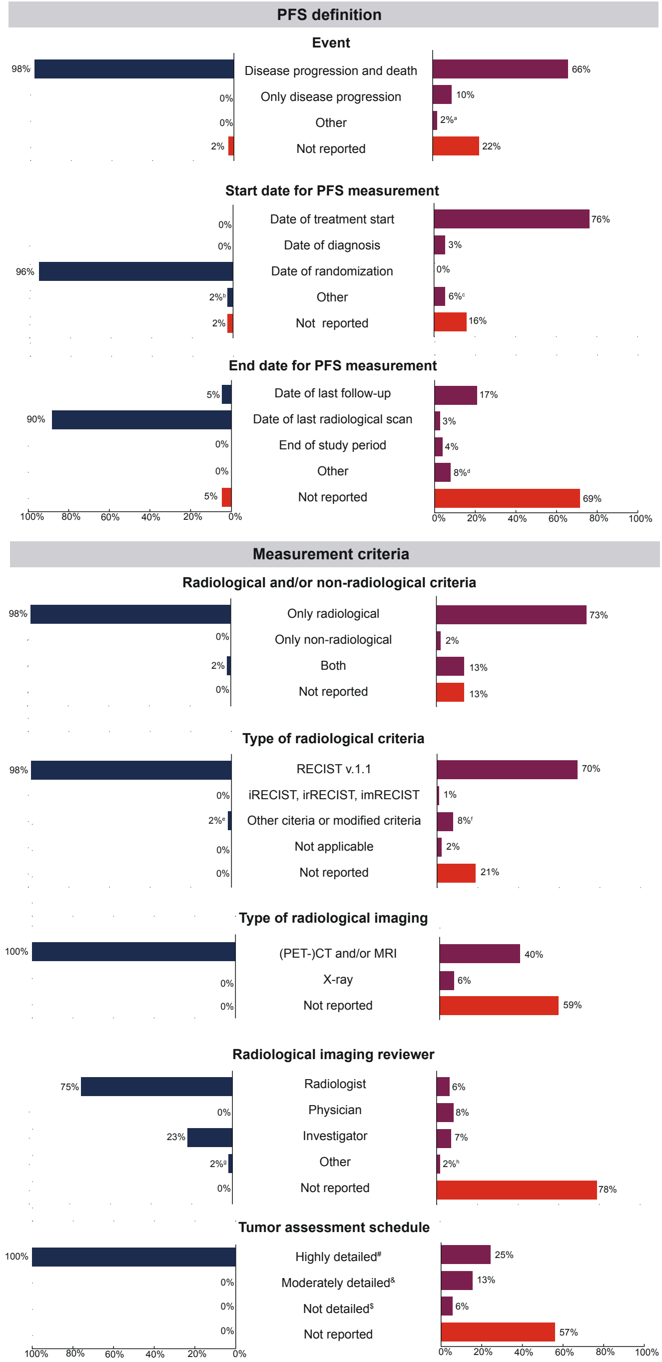


**Figure 2.** Horizontal bar chart displaying PFS measurement characteristics of the included clinical trials and observational studies

Abbreviations: iRECIST, immune-related response evaluation criteria of solid tumors; irRECIST, immunotherapy response evaluation criteria of solid tumors; imRECIST, immune-modified response evaluation criteria of

solid tumors; RECIST, response evaluation criteria of solid tumors; (PET-)CT, (positron emission tomography-) computed tomograpgy; PFS, progression-free survival; MRI, magnetic resonance imaging.
± Percentages are calculated by dividing the number of each category of 'type of radiological criteria' by the total number of studies that used radiological criteria.
# Highly detailed response assessment schedules are expressed in exact units such as cycles, days, weeks or months with a no significant allowance for variation (+/- 7 days or less). For example, response assessments occur every 6 weeks.
& Moderately detailed response assessment schedules are expressed in units such as cycles, days, weeks or months with a significant allowance for variation (>7 days). For example, response assessments occur approximately every 2 to 3 months.
$ Not detailed response assessment schedules lacking precise timeframes. For example, response assessments occur at the physician's discretion.
a Three observational studies used other event definitions: disease progression, death or change of treatment (n = 1), disease progression or change of treatment (n = 1), and tumor relapse or death (n=1).
b One clinical trial used the date of sub-study registration as start date of PFS evaluation.
c Nine observational studies used other start dates for PFS evaluation: the last date of (chemo)radiotherapy (n = 7), the last date of radiotherapy plus extra days to correct for immortal time bias (n=1), and the first day of no evidence of disease progression (n=1).
d Nine observational studies used other end dates for PFS evaluation: the last contact date or cut-off date if earlier (n = 6), and the last contact date of start date of new treatment line (n = 1).
e One clinical trial used the mWHO criteria to detect disease progression.
f Eleven observational studies used other or modified radiological criteria: RECIST modified to detect pseudoprogression (n = 4), modified RECIST (n = 5), immune-related response criteria
(n = 1), and RANO criteria (n = 1).
g In one clinical trial, radiological imaging was reviewed by a combination of a qualified study physician and a radiologist to review radiological imaging.
h Four observational studies used other radiological imaging reviewers: a combination of a radiologist and a physician (n = 2), and a combination of an investigator and a radiologist (n = 1).

## Discussion

This study was the first to systematically evaluate how PFS was defined, measured and reported across RCTs and observational studies. The results show that, in contrast to RCTs, observational studies often lack reporting on PFS definitions and measurements, with 69% not reporting the end date for PFS measurement, 59% not reporting the type of radiological imaging, 78% not reporting the imaging reviewer, and 57% not reporting the tumor assessment schedule. For studies that did report on PFS definitions and measurements, these often differed from those in RCTs, particularly regarding event definitions, the start- and stop dates for PFS measurement, and tumor assessment schedules. These findings hinder the interpretability and reliability of PFS outcomes of observational studies, limiting their ability to complement evidence from RCTs.

The low level of reporting of PFS measurement characteristics in observational studies included lacking information on the end date of PFS calculation, the type of radiological imaging and its reviewer, and the tumor assessment schedule. This lack of information can lead to unreliable estimates of PFS that are difficult to interpret and unreliable to use in clinical practice, undermining clinicians' ability to make informed treatment choices. At the time these studies were conducted, general reporting standards for observational research such as STROBE [9] and RECORD-PE [10], were available and included recommendations for clearly defining outcomes. However, specific guidance to oncological observational studies, such as those later provided by the ESMO [12] and ASCO [13], were not available, which may have contributed to the insufficient reporting. Although the introduction of these guidance documents is a step forward, they still do not provide specific instructions on reporting PFS measurement characteristics. Detailed reporting instructions on definitions of PFS, including events and follow-up times, and measurement criteria for PFS, including radiological criteria, assessment intervals, radiological imaging and reviewers, are needed to improve the interpretability and reliability of PFS estimates from observational studies.

The variability in PFS definitions can be attributed to both fundamental flaws in study design choices and inherent limitations in the design of observational studies. One such fundamental flaw is using only disease progression as relevant event in PFS estimates, which could lead to biased estimates of PFS. An inherent limitation of observational studies is that PFS is measured from the date of treatment start, whereas in RCTs PFS is measured from the date of randomization. Also, the observed variability in PFS measurement criteria could result from inherent limitations of the observational studies since they leverage data from real-world clinical practices rather than highly standardized settings. For example, response assessments in observational studies occur less frequently and consistently than in RCTs, causing delayed or missed detection of disease progression, introducing assessment bias in PFS comparisons with RCTs as well as other observational studies [5,21]. While some of the variabilities in PFS measurement criteria, such as the choice of imaging modalities or the frequency of response assessments, may be unavoidable due to the practical constraints of routine clinical care, others can be minimized through high-quality study design and methodologies. A recommended approach is to use the target trial emulation framework, which helps to design observational studies by closely replicating the design of hypothetical RCTs [22]. Aligning key elements of trials, including strict PFS definitions, radiological criteria, imaging reviewers, and - where feasible- assessment schedules, reduces the variability in PFS measurement characteristics between observational studies and between observational studies and RCTs, thereby reducing bias in PFS comparisons between studies.

Some inherent variability in PFS measurement characteristics between observational studies and RCTs will always remain, which limits the ability to use evidence from observational studies to complement evidence from RCTs. In clinical practice, physicians evaluate disease progression to guide individual treatment decisions, resulting in less standardized measurement of PFS compared to RCTs. In clinical practice, the evaluation of disease progression is also based on clinical criteria such as worsening of clinical symptoms and performance status rather than solely on RECIST-evaluated radiological imaging.[25] Additionally, the timing and use of scans in clinical practice are based on clinical needs and available resources rather than strict predefined protocols. In contrast, RCTs only use RECIST-evaluated radiological imaging and strict assessment schedules to evaluate disease progression. These variations in evaluation methods can result in biased PFS outcomes. [5] [26]This can be illustrated by the recent findings from the PACIFIC-R study [23], an observational study designed to complement the PACIFIC RCT [24] by evaluating the effectiveness of adjuvant durvalumab in patients with unresectable NSCLC. The PACIFIC-R study reported a longer median PFS compared to the PACIFIC RCT (24.9 versus 16.9 months). Similarly, findings from our previous study on durvalumab showed longer PFS in patients treated in clinical practice compared to those in the PACIFIC RCT (27.1 versus 16.9 months) [27]. Both examples illustrate that without a standardized approach, PFS from observational studies may not capture the same clinical outcome as PFS from RCTs, raising the question whether PFS outcomes in observational studies can be reliably used to complement PFS outcomes from RCTs. Using the term 'real-world PFS' for PFS outcomes in observation studies may help to highlight the difference in measurement and context, as also discussed by others. [28, 29] Other endpoints, such as time to treatment discontinuation and time to treatment failure, have been proposed as more pragmatic alternatives for assessing treatment effects in clinical practice, although they also have their limitations [30]. Nevertheless, PFS estimates from observational studies are valuable in situations where there is a total lack of evidence from RCTs. For example, they provide valuable insights into specific subpopulations not represented in RCTs and into treatment comparisons where no head-to-head RCT is available. While PFS estimates from observational studies are fit-for-purpose in these situations, they should be evaluated critically due to the inherent variability in their measurement.

Several limitations of our scoping review should be considered. First, the review was limited in scope as it focused primarily on immunotherapy treatments for advanced NSCLC, which may not fully represent the broader range of issues across different types of cancer and treatment modalities. However, we believe that the observed variation in PFS measurement characteristics and lack of reporting is likely to apply to other settings as well. Additionally, while there is a small risk of missing studies due to the AI-assisted review tool, the review of nearly 10000 studies support the reliability of our findings, even if a few studies were missed.

## Conclusion

The findings from our scoping review show that, in contrast to RCTs, observational studies often lack reporting on PFS definitions and measurements. When these characteristics are reported, they often differ across observational studies and between them and RCTs. Since observational studies are important for complementing evidence, aligning PFS definition and measurement criteria with those used in RCT, along with detailed reporting on these aspects, is needed. However, some variability in PFS measurement characteristics is unavoidable, and therefore, PFS estimates from observational studies should be interpreted critically and carefully.

## References

1. Del Paggio JC, Berry JS, Hopman WM, Eisenhauer EA, Prasad V, Gyawali B, et al. Evolution of the randomized clinical trial in the era of precision oncology. JAMA Oncol. 2021 May 1;7(5):728–34
2. Tang M, Pearson SA, Simes RJ, Chua BH. Harnessing Real-World Evidence to Advance Cancer Research. Vol. 30, Current Oncology. MDPI; 2023. p. 1844–59.
3. Nordon C, Karcher H, Groenwold RHH, Ankarfeldt MZ, Pichler F, Chevrou-Severac H, et al. The “efficacy-effectiveness gap”: Historical background and current conceptualization. Value in Health. 2016;19(1):75–81.
4. Korn RL, Crowley JJ. Overview: Progression-free survival as an endpoint in clinical trials with solid tumors. Clinical Cancer Research. 2013;19(10):2607–12.
5. Panageas KS, Ben-Porat L, Dickler MN, Chapman PB, Schrag D. When you look matters: The effect of assessment schedule on progression-free survival. J Natl Cancer Inst. 2007;99(6):428–32.
6. Sullivan DC, Schwartz LH, Zhao B. The imaging viewpoint: How imaging affects determination of progression-free survival. Vol. 19, Clinical Cancer Research. American Association for Cancer Research Inc.; 2013. p. 2621–8.
7. Adamson BJS, Ma X, Griffith SD, Sweeney EM, Sarkar S, Bourla AB. Differential frequency in imaging-based outcome measurement: Bias in real-world oncology comparative-effectiveness studies. Pharmacoepidemiol Drug Saf. 2022;31(1):46–54.
8. Booth CM, Eisenhauer EA. Progression-free survival: Meaningful or simply measurable? Journal of Clinical Oncology. 2012;30(10):1030–3.
9. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol. 2008 Apr;61(4):344–9.
10. Langan SM, Schmidt SA, Wing K, Ehrenstein V, Nicholls SG, Filion KB, et al. The reporting of studies conducted using observational routinely collected health data statement for pharmacoepidemiology (RECORD-PE). Vol. 363, The BMJ. BMJ Publishing Group; 2018.
11. Castelo-Branco L, Pellat A, Martins-Branco D, Valachis A, Derksen JWG, Suijkerbuijk KPM, et al. ESMO Guidance for Reporting Oncology real-World evidence (GROW). ESMO Real World Data and Digital Oncology. 2023 Nov;1:100003.
12. Ramsey SD, Onar-Thomas A, Wheeler SB. Real-World Database Studies in Oncology: A Call for Standards. Vol. 42, Journal of Clinical Oncology. Lippincott Williams and Wilkins; 2024. p. 977–80.
13. Peters M, Godfrey CM, Mcinerney P, Soares CB. Methodology for JBI Scoping Reviews. Int J Evid Based Healthc. 2015 Sep;13(3):141-6
14. Tricco AC, Lillie E, Zarin W, O’Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. Vol. 169, Annals of Internal Medicine. American College of Physicians; 2018. p. 467–73.
15. van de Schoot R, de Bruin J, Schram R, Zahedi P, de Boer J, Weijdema F, et al. An open source machine learning framework for efficient and transparent systematic reviews. Nat Mach Intell. 2021 Feb 1;3(2):125–33.
16. ASReview: Active learning for systematic reviews. How to stop screening. [cited on 23 April 2024]. Available from: https://github.com/asreview/asreview/discussions/557
17. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). Eur J Cancer. 2009 Jan;45(2):228–47.
18. Nishino M, Giobbie-Hurder A, Gargano M, Suda M, Ramaiya NH, Hodi FS. Developing a common language for tumor response to immunotherapy: Immune-related response criteria using unidimensional measurements. Clinical Cancer Research. 2013 Jul 15;19(14):3936–43.

19. Seymour L, Bogaerts J, Perrone A, Ford R, Schwartz LH, Mandrekar S, et al. iRECIST: guidelines for response criteria for use in trials testing immunotherapeutics. Lancet Oncol. 2017;18(3):143–52
20. Stephen Hodi F, Ballinger M, Lyons B, Soria JC, Nishino M, Tabernero J, et al. Immune-modified response evaluation criteria in solid tumors (imrecist): Refining guidelines to assess the clinical benefit of cancer immunotherapy. Journal of Clinical Oncology. 2018 Jan 17;36(9):850–8.
21. Haslam A, Gill J, Prasad V. The frequency of assessment of progression in randomized oncology clinical trials. Cancer Rep. 2022 Jul 1;5(7).
22. Matthews AA, Danaei G, Islam N, Kurth T. Target trial emulation: Applying principles of randomised trials to observational studies. The BMJ. 2022;
23. Filippi AR, Bar J, Chouaid C, Christoph DC, Field JK, Fietkau R, et al. Real-world outcomes with durvalumab after chemoradiotherapy in patients with unresectable stage III NSCLC: interim analysis of overall survival from PACIFIC-R. ESMO Open. 2024 Jun 1;9(6).
24. Spigel DR, Faivre-Finn C, Gray JE, Vicente D, Planchard D, Paz-Ares L, et al. Five-Year Survival Outcomes From the PACIFIC Trial: Durvalumab After Chemoradiotherapy in Stage III Non–Small-Cell Lung Cancer. Journal of Clinical Oncology. 2022;
25. Griffth SD, Tucker M, Bowser B, Calkins G, Joe Chang CH, Guardino E, et al. Generating real-world tumor burden endpoints from electronic health record data: Comparison of RECIST, radiology-anchored, and clinician-anchored approaches for abstracting real-world progression in Non-small cell lung cancer. Adv Ther. 2019 Aug.36(8).2122-2136
26. Amorrotu R, Garcia M, Zhao Y, Naqa IE, Balagurunathan Y, Chen DT, et al. Overview of approaches to estimate real-world disease progression in lung cancer. JNCI Cancer Spectrum. 2023 Oct. 7(6).
27. Verschueren M V, Cg A. Durvalumab after chemoradiotherapy in patients with stage III non-small-cell lung cancer : real-world outcomes versus clinical trial results. 2023;
28. Mahtre SK, Machado RJK, Ton TGN, Trinh H, Mazieres J, Rittmeyer A, et al. Real-world progression-free survival as an endpoint in lung cancer: Replicating atezolizumab and docetaxel arms of the OAK trial using real-world data. Clin Pharmacol Ther. 2023 dec.114(6).1313-1322.
29. Torres A, Nussbaum N, Parrinello C, Bourla A, Bowser B, Wagner S, et al. Analysis of a real-world progression variable and related endpoints in patients with five different cancer types. 2022. 39(6). 2832-2849.
30. Blumenthal GM, Gong Y, Kehl K, Mishra-Kalyani P, Goldberg KB, Khozin S, 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. Annals of Oncology. 2019 May 1;30(5):830–8.

## RCTs

31. Antonia SJ, Villegas A, Daniel D, Vicente D, Murakami S, Hui R, et al. Durvalumab after Chemoradiotherapy in Stage III Non–Small-Cell Lung Cancer. New England Journal of Medicine. 2017 Nov 16;377(20):1919–29.
32. Barlesi F, Vansteenkiste J, Spigel D, Ishii H, Garassino M, de Marinis F, et al. Avelumab versus docetaxel in patients with platinum-treated advanced non-small-cell lung cancer (JAVELIN Lung 200): an open-label, randomised, phase 3 study. Lancet Oncol. 2018 Nov 1;19(11):1468–79.
33. Brahmer J, Reckamp KL, Baas P, Crinò L, Eberhardt WEE, Poddubskaya E, et al. Nivolumab versus Docetaxel in Advanced Squamous-Cell Non–Small-Cell Lung Cancer. New England Journal of Medicine. 2015 Jul 9;373(2):123–35.

34. Borghaei H, Paz-Ares L, Horn L, Spigel DR, Steins M, Ready NE, et al. Nivolumab versus Docetaxel in Advanced Nonsquamous Non–Small-Cell Lung Cancer. New England Journal of Medicine. 2015 Oct 22;373(17):1627–39.
35. Boyer M, N S MA, Rodríguez-Abreu D, Park K, Ho Lee D, Çiçin I, et al. Pembrolizumab Plus Ipilimumab or Placebo for Metastatic Non-Small-Cell Lung Cancer With PD-L1 Tumor Proportion Score ≥ 50%: Randomized, Double-Blind Phase III KEYNOTE-598 Study. J Clin Oncol.2021;39:2327–38.
36. Carbone DP, Reck M, Paz-Ares L, Creelan B, Horn L, Steins M, et al. First-Line Nivolumab in Stage IV or Recurrent Non–Small-Cell Lung Cancer. New England Journal of Medicine. 2017 Jun 22;376(25):2415–26.
37. de Castro G, Rizvi NA, Schmid P, Syrigos K, Martin C, Yamamoto N, et al. NEPTUNE: Phase 3 Study of First-Line Durvalumab Plus Tremelimumab in Patients With Metastatic NSCLC. Journal of Thoracic Oncology. 2023 Jan 1;18(1):106–19.
38. Gandhi L, Rodríguez-Abreu D, Gadgeel S, Esteban E, Felip E, De Angelis F, et al. Pembrolizumab plus Chemotherapy in Metastatic Non–Small-Cell Lung Cancer. New England Journal of Medicine. 2018 May 31;378(22):2078–92.
39. Gettinger SN, Redman MW, Bazhenova L, Hirsch FR, Mack PC, Schwartz LH, et al. Nivolumab plus Ipilimumab vs Nivolumab for Previously Treated Patients with Stage IV Squamous Cell Lung Cancer: The Lung-MAP S1400I Phase 3 Randomized Clinical Trial. JAMA Oncol. 2021 Sep 1;7(9):1368–77.
40. Gogishvili M, Melkadze T, Makharadze T, Giorgadze D, Dvorkin M, Penkov K, et al. Cemiplimab plus chemotherapy versus chemotherapy alone in non-small cell lung cancer: a randomized, controlled, double-blind phase 3 trial. Nat Med. 2022 Nov 1;28(11):2374–80.
41. Govindan R, Szczesna A, Ahn MJ, Schneider CP, Fernando Gonzalez Mella P, Barlesi F, et al. JOURNAL OF CLINICAL ONCOLOGY Phase III Trial of Ipilimumab Combined With Paclitaxel and Carboplatin in Advanced Squamous Non-Small-Cell Lung Cancer. J Clin Oncol. 2017;35:3449–57.
42. Johnson ML, Chul Cho B, Luft A, Alatorre-Alexander J, Lucien Geater S, Laktionov K, et al. Durvalumab With or Without Tremelimumab in Combination With Chemotherapy as First-Line Therapy for Metastatic Non-Small-Cell Lung Cancer: The Phase III POSEIDON Study. J Clin Oncol. 2022;41:1213–27.
43. Hellmann MD, Ciuleanu TE, Pluzanski A, Lee JS, Otterson GA, Audigier-Valette C, et al. Nivolumab plus Ipilimumab in Lung Cancer with a High Tumor Mutational Burden. New England Journal of Medicine. 2018 May 31;378(22):2093–104.
44. Herbst RS, Giaccone G, de Marinis F, Reinmuth N, Vergnenegre A, Barrios CH, et al. Atezolizumab for First-Line Treatment of PD-L1–Selected Patients with NSCLC. New England Journal of Medicine. 2020 Oct 1;383(14):1328–39.
45. Jotte R, Cappuzzo F, Vynnychenko I, Stroyakovskiy D, Rodríguez-Abreu D, Hussein M, et al. Atezolizumab in Combination With Carboplatin and Nab-Paclitaxel in Advanced Squamous NSCLC (IMpower131): Results From a Randomized Phase III Trial. Journal of Thoracic Oncology. 2020 Aug 1;15(8):1351–60.
46. Nishio M, Barlesi F, West H, Ball S, Bordoni R, Cobo M, et al. Atezolizumab Plus Chemotherapy for First-Line Treatment of Nonsquamous NSCLC: Results From the Randomized Phase 3 IMpower132 Trial. Journal of Thoracic Oncology. 2021 Apr 1;16(4):653–64.
47. Paz-Ares L, Ciuleanu TE, Cobo M, Schenker M, Zurawski B, Menezes J, et al. First-line nivolumab plus ipilimumab combined with two cycles of chemotherapy in patients with non-small-cell lung cancer (CheckMate 9LA): an international, randomised, open-label, phase 3 trial. Lancet Oncol. 2021 Feb 1;22(2):198–211.

48. Paz-Ares L, Vicente D, Tafreshi A, Robinson A, Soto Parra H, Mazières J, et al. A Randomized, Placebo-Controlled Trial of Pembrolizumab Plus Chemotherapy in Patients With Metastatic Squamous NSCLC: Protocol-Specified Final Analysis of KEYNOTE-407. Journal of Thoracic Oncology. 2020;15(10):1657–69.
49. Peters S, Dziadziuszko R, Morabito A, Felip E, Gadgeel SM, Cheema P, et al. Atezolizumab versus chemotherapy in advanced or metastatic NSCLC with high blood-based tumor mutational burden: primary analysis of BFAST cohort C randomized phase 3 trial. Nat Med. 2022 Sep 1;28(9):1831–9.
50. Planchard D, Reinmuth N, Orlov S, Fischer JR, Sugawara S, Mandziuk S, et al. ARCTIC: durvalumab with or without tremelimumab as third-line or later treatment of metastatic non-small-cell lung cancer. Annals of Oncology. 2020 May 1;31(5):609–18.
51. Mok TSK, Wu YL, Kudaba I, Kowalski DM, Cho BC, Turna HZ, et al. Pembrolizumab versus chemotherapy for previously untreated, PD-L1-expressing, locally advanced or metastatic non-small-cell lung cancer (KEYNOTE-042): a randomized, open-label, controlled, phase 3 trial. The Lancet. 2019 May 4;393(10183):1819–30.
52. Reck M, Barlesi F, Yang JCH, Westeel V, Felip E, Özgüroğlu M, et al. Avelumab Versus Platinum-Based Doublet Chemotherapy as First-Line Treatment for Patients With High-Expression Programmed Death-Ligand 1–Positive Metastatic NSCLC: Primary Analysis From the Phase 3 JAVELIN Lung 100 Trial. Journal of Thoracic Oncology. 2023;
53. Reck M, Rodríguez-Abreu D, Robinson AG, Hui R, Csőszi T, Fülöp A, et al. Pembrolizumab versus Chemotherapy for PD-L1–Positive Non–Small-Cell Lung Cancer. New England Journal of Medicine. 2016 Nov 10;375(19):1823–33.
54. Rittmeyer A, Barlesi F, Waterkamp D, Park K, Ciardiello F, von Pawel J, et al. Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (OAK): a phase 3, open-label, multicentre randomised controlled trial. The Lancet. 2017 Jan 21;389(10066):255–65.
55. Rizvi NA, Cho BC, Reinmuth N, Lee KH, Luft A, Ahn MJ, et al. Durvalumab with or Without Tremelimumab vs Standard Chemotherapy in First-line Treatment of Metastatic Non-Small Cell Lung Cancer: The MYSTIC Phase 3 Randomized Clinical Trial. JAMA Oncol. 2020 May 1;6(5):661–74.
56. 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. 2021. 397.
57. Shi Y, Wu L, Yu X, Xing P, Wang Y, Zhou J, et al. Sintilimab versus docetaxel as second-line treatment in advanced or metastatic squamous non-small-cell lung cancer: an open-label, randomized controlled phase 3 trial (ORIENT-3). Cancer Commun. 2022 Dec 1;42(12):1314–30.
58. Socinski MA, Jotte RM, Cappuzzo F, Orlandi F, Stroyakovskiy D, Nogami N, et al. Atezolizumab for First-Line Treatment of Metastatic Nonsquamous NSCLC. New England Journal of Medicine. 2018 Jun 14;378(24):2288–301
59. Sugawara S, Lee JS, Kang JH, Kim HR, Inui N, Hida T, et al. Nivolumab with carboplatin, paclitaxel, and bevacizumab for first-line treatment of advanced nonsquamous non-small-cell lung cancer. Annals of Oncology. 2021 Sep 1;32(9):1137–47.
60. Wang J, Lu S, Yu X, Hu Y, Sun Y, Wang Z, et al. Tislelizumab plus chemotherapy vs chemotherapy alone as first-line treatment for advanced squamous non-small-cell lung cancer a phase 3 randomized clinical trial. JAMA Oncol. 2021 May 1;7(5):709–17.
61. Wang Z, Wu L, Li B, Cheng Y, Li X, Wang X, et al. Toripalimab Plus Chemotherapy for Patients With Treatment-Naive Advanced Non-Small-Cell Lung Cancer: A Multicenter Randomized Phase III Trial (CHOICE-01). J Clin Oncol. 2022;41:651–63.

62. West H, McCleod M, Hussein M, Morabito A, Rittmeyer A, Conter HJ, et al. Atezolizumab in combination with carboplatin plus nab-paclitaxel chemotherapy compared with chemotherapy alone as first-line treatment for metastatic non-squamous non-small-cell lung cancer (IMpower130): a multicentre, randomised, open-label, phase 3 trial. Lancet Oncol. 2019 Jul 1;20(7):924–37.
63. Wu YL, Lu S, Cheng Y, Zhou C, Wang J, Mok T, et al. Nivolumab Versus Docetaxel in a Predominantly Chinese Patient Population With Previously Treated Advanced NSCLC: CheckMate 078 Randomized Phase III Clinical Trial. Journal of Thoracic Oncology. 2019 May 1;14(5):867–75.
64. Yang JCH, Han B, De La Mora Jiménez E, Lee JS, Koralewski P, Karadurmus N, et al. Pembrolizumab With or Without Lenvatinib for First-Line Metastatic Non-Small-Cell Lung Cancer With PD-L1 Tumor Proportion Score ≥1% (LEAP-007): A Randomized, Double-Blind, Phase 3 Trial. Journal of Thoracic Oncology. 2023 Dec;
65. Wu YL, Lu S, Cheng Y, Zhou C, Wang J, Mok T, et al. Nivolumab Versus Docetaxel in Predominantly Chinese Patient Population With Previously Treated Advanced NSCLC: CheckMate 078 Randomized Phase III Clinical Trial. Journal of Thoracic Oncology. 2019 May 1;14(5):867–75.
66. Zhang L, Wang Z, Fang J, Yu Q, Han B, Cang S, et al. Final overall survival data of sintilimab plus pemetrexed and platinum as First-Line treatment for locally advanced or metastatic nonsquamous NSCLC in the Phase 3 ORIENT-11 study. Lung Cancer. 2022 Sep 1;171:56–60.
67. Zhou C, Chen G, Huang Y, Zhou J, Lin LZ, Feng J, et al. Camrelizumab plus carboplatin and pemetrexed versus chemotherapy alone in chemotherapy-naive patients with advanced non-squamous non-small-cell lung cancer (CameL): a randomised, open-label, multicentre, phase 3 trial. Lancet Respir Med. 2021 Mar 1;9(3):305–14.
68. Zhou Q, Chen M, Jiang O, Pan Y, Hu D, Lin Q, et al. Sugemalimab versus placebo after concurrent or sequential chemoradiotherapy in patients with locally advanced, unresectable, stage III non-small-cell lung cancer in China (GEMSTONE-301): interim results of a randomised, double-blind, multicentre, phase 3 trial. Lancet Oncol. 2022 Feb 1;23(2):209–19.
69. Zhou C, Wang Z, Sun Y, Cao L, Ma Z, Wu R, et al. Sugemalimab versus placebo, in combination with platinum-based chemotherapy, as first-line treatment of metastatic non-small-cell lung cancer (GEMSTONE-302): interim and final analyses of a double-blind, randomised, phase 3 clinical trial. Lancet Oncol. 2022 Feb 1;23(2):220–33
70. Zhou C, Wu L, Fan Y, Wang Z, Liu L, Chen G, et al. Sintilimab Plus Platinum and Gemcitabine as First-Line Treatment for Advanced or Metastatic Squamous NSCLC: Results From a Randomized, Double-Blind, Phase 3 Trial (ORIENT-12). Journal of Thoracic Oncology. 2021 Sep 1;16(9):1501–11.

## Observational studies

71. Abe T, Saito S, Iino M, Aoshika T, Ryuno Y, Ohta T, et al. Effect of durvalumab on local control after concurrent chemoradiotherapy for locally advanced non-small cell lung cancer in comparison with chemoradiotherapy alone. Thorac Cancer. 2021 Jan 1;12(2):245–50.
72. Abe T, Iino M, Saito S, Aoshika T, Ryuno Y, Ohta T, et al. Comparison of the Efficacy and Toxicity of Concurrent Chemoradiotherapy and Durvalumab and Concurrent Chemoradiotherapy Alone for Locally Advanced Non-small Cell Lung Cancer With N3 Lymph Node Metastasis. Anticancer Res. 2023 Feb 1;43(2):675–82.
73. Afzal MZ, Dragnev K, Shirai K. A tertiary care cancer center experience with carboplatin and pemetrexed in combination with pembrolizumab in comparison with carboplatin and pemetrexed alone in non-squamous non-small cell lung cancer. J Thorac Dis. 2018 Jun 1;10(6):3575–84.

74. Ahn BC, Pyo KH, Xin CF, Jung D, Shim HS, Lee CY, et al. Comprehensive analysis of the characteristics and treatment outcomes of patients with non-small cell lung cancer treated with anti-PD-1 therapy in real-world practice. J Cancer Res Clin Oncol. 2019;
75. Merino Almazán M, Duarte Pérez JM, Marín Pozo JF, Ortega Granados AL, Muros De Fuentes B, Quesada Sanz P, et al. A multicentre observational study of the effectiveness, safety and economic impact of nivolumab on non-small-cell lung cancer in real clinical practice. Int J Clin Pharm. 2019 Feb 15;41(1):272–9.
76. Alonso-García M, Sánchez-Gastaldo A, Muñoz-Fuentes MA, Molina-Pinelo S, Boyero L, Benedetti JC, et al. Real-World Analysis of Nivolumab and Atezolizumab Efficacy in Previously Treated Patients with Advanced Non-Small Cell Lung Cancer. Pharmaceuticals (Basel). 2022 May 1 [cited 2022 Jul 28];15(5).
77. Amrane K, Geier M, Corre R, Léna H, Léveiller G, Gadby F, et al. First-line pembrolizumab for non–small cell lung cancer patients with PD-L1 ≥50% in a multicenter real-life cohort: The PEMBREIZH study. Cancer Med. 2020 Apr 1;9(7):2309–16.
78. Attili I, Valenza C, Santoro C, Antonarelli G, Trillo Aliaga P, Del Signore E, et al. Comparison of real-world data (RWD) analysis on efficacy and post-progression outcomes with pembrolizumab plus chemo vs chemo alone in metastatic non-squamous non-small cell lung cancer with PD-L1 < 50%. Front Oncol. 2022 Aug 10;12.
79. Batra U, Chufal KS, Nathany S, Ahmad I, Chowdhary RL, Sharma M, et al. Immunotherapy in advanced non-small-cell lung cancer (NSCLC) after progression on chemotherapy: Real-world results from a prospective institutional cohort. Immunotherapy. 2022 Aug 1;14(11):851–8.
80. Bazhenova L, Kish J, Cai B, Caro N, Feinberg B. Real-world observational study of current treatment patterns and outcomes in recurrent or locally advanced/metastatic non-small cell lung cancer. Cancer Treat Res Commun. 2022 Jan 1;33.
81. Benjamin DJ, Chen S, Eldredge JB, Schokrpur S, Li D, Quan Z, et al. The Role of Chemotherapy Plus Immune Checkpoint Inhibitors in Oncogenic-Driven NSCLC: A University of California Lung Cancer Consortium Retrospective Study. JTO Clin Res Rep. 2022 Dec 1;3(12).
82. Bjørnhart B, Hansen KH, Jørgensen TL, Herrstedt J, Schytte T. Efficacy and safety of immune checkpoint inhibitors in a Danish real life non-small cell lung cancer population: a retrospective cohort study. Acta Oncol (Madr). 2019 Jul 3;58(7):953–61.
83. Bongiovanni A, Foca F, Menis J, Stucci SL, Artioli F, Guadalupi V, et al. Immune Checkpoint Inhibitors With or Without Bone-Targeted Therapy in NSCLC Patients With Bone Metastases and Prognostic Significance of Neutrophil-to-Lymphocyte Ratio. Front Immunol. 2021 Nov 10;12.
84. Bruni A, Scotti V, Borghetti P, Vagge S, Cozzi S, D'Angelo E, et al. A Real-World, Multicenter, Observational Retrospective Study of Durvalumab After Concomitant or Sequential Chemoradiation for Unresectable Stage III Non-Small Cell Lung Cancer. Front Oncol. 2021 Sep 28;11.
85. Chen Y, Wang Y, Yang Z, Hu M, Zhang Y, Qian F, et al. Pembrolizumab Alone or Combined With Chemotherapy in Advanced NSCLC With PD-L1 ≥50%: Results of a Retrospective Study. Front Oncol. 2021 Jun 28;11.
86. José ABS, Colomer-Aguilar C, Martínez-Caballero D, Massutí-Sureda B. Effectiveness and safety of atezolizumab, nivolumab and pembrolizumab in metastatic non-small cell lung cancer. In: Farmacia Hospitalaria. Grupo Aula Medica S.L.; 2021. p. 121–5.
87. Calpe-Armero P, Ferriols-Lisart R, Ferriols-Lisart F, Pérez-Pitarch A. Effectiveness of Nivolumab versus Docetaxel as Second-Line Treatment in Non-Small Cell Lung Cancer Patients in Clinical Practice. Chemotherapy. 2017 Nov 1;62(6):374–80.

88. Cavaille F, Peretti M, Garcia ME, Giorgi R, Ausias N, Vanelle P, et al. Real-world efficacy and safety of pembrolizumab in patients with non-small cell lung cancer: a retrospective observational study. Tumori. 2021 Feb 1;107(1):32–8.
89. Chen B, Wang J, Pu X, Li J, Wang Q, Liu L, et al. The efficacy and safety of immune checkpoint inhibitors combined with chemotherapy or anti-angiogenic therapy as a second-line or later treatment option for advanced non-small cell lung cancer: a retrospective comparative cohort study. Transl Lung Cancer Res. 2022 Oct 1;11(10):2111–24.
90. Chen Y, Kang S, Yan M. Atezolizumab plus carboplatin and nab-paclitaxel versus carboplatin and nab-paclitaxel as treatments for Chinese, treatment-naïve, stage IV, non-squamous, non-small-cell lung cancer patients: A retrospective analysis. Pharmacol Res Perspect. 2022 Jun 1;10(3).
91. Chen S, Wei H, Zhao W, Jiang W, Ning R, Zhou S, et al. PD-1/PD-L1 inhibitors plus anti-angiogenic agents with or without chemotherapy versus PD-1/PD-L1 inhibitors plus chemotherapy as second or later-line treatment for patients with advanced non-small cell lung cancer: A real-world retrospective cohort study. Front Immunol. 2022 Dec 7;13.
92. Chen M, Xu Y, Zhao J, Liu X, Liu X, Zhang D, et al. Comparison of Chemotherapy Plus Pembrolizumab vs. Chemotherapy Alone in EGFR-Mutant Non–small-Cell Lung Cancer Patients. Clin Lung Cancer. 2023 May 1;24(3):278–86.
93. Chen Y, Yang Z, Wang Y, Hu M, Zhang B, Zhang Y, et al. Pembrolizumab Plus Chemotherapy or Anlotinib vs. Pembrolizumab Alone in Patients With Previously Treated EGFR-Mutant NSCLC. Front Oncol. 2021 Apr 16;11.
94. Cheng Y, Yang B, Ouyang W, Jie C, Zhang W, Chen G, et al. Is ICI-based therapy better than chemotherapy for metastatic NSCLC patients who develop EGFR-TKI resistance? A real-world investigation. Front Oncol. 2022 Aug 23;12.
95. Chu X, Qiang H, Xie M, Li X, Zhao J, Wu Y, et al. Treatment efficacy of HER2-mutant lung adenocarcinoma by immune checkpoint inhibitors: a multicenter retrospective study. Cancer Immunology, Immunotherapy. 2022 Jul 1;71(7):1625–31.
96. Cramer-van der Welle CM, Verschueren M V., Tonn M, Peters BJM, Schramel FMNH, Klungel OH, et al. Real-world outcomes versus clinical trial results of immunotherapy in stage IV non-small cell lung cancer (NSCLC) in the Netherlands. Sci Rep. 2021 Dec 1;11(1).
97. Cui P, Li R, Huang Z, Wu Z, Tao H, Zhang S, et al. Comparative effectiveness of pembrolizumab vs. nivolumab in patients with recurrent or advanced NSCLC. Sci Rep. 2020 Dec 1;10(1).
98. Deng H, Lin X, Xie X, Yang Y, Wang L, Wu J, et al. Immune Checkpoint Inhibitors Plus Single-Agent Chemotherapy for Advanced Non-Small-Cell Lung Cancer After Resistance to EGFR-TKI. Front Oncol. 2021 Sep 20;11.
99. Descourt R, Greillier L, Perol M, Ricordel C, Auliac JB, Falchero L, et al. First-line single-agent pembrolizumab for PD-L1-positive (tumor proportion score ≥ 50%) advanced non-small cell lung cancer in the real world: impact in brain metastasis: a national French multicentric cohort (ESCKEYP GFPC study). Cancer Immunology, Immunotherapy. 2023 Jan 1;72(1):91–9.
100. Diker O, Olgun P. First-line pembrolizumab efficacy in patients with advanced non-small cell lung cancer: A Bi-center retrospective, real-life experience study. JBUON. 2021;26(3):844–52.
101. Dudnik E, Moskovitz M, Daher S, Shamai S, Hanovich E, Grubstein A, et al. Effectiveness and safety of nivolumab in advanced non-small cell lung cancer: The real-life data. Lung Cancer. 2018 Dec 1;126:217–23.
102. Faehling M, Schumann C, Christopoulos P, Hoffknecht P, Alt J, Horn M, et al. Durvalumab after definitive chemoradiotherapy in locally advanced unresectable non-small cell lung cancer (NSCLC): Real-world data on survival and safety from the German expanded-access program (EAP). Lung Cancer. 2020 Dec 1;150:114–22.

103. Feng Y, Tang L, Wang H, Liu Y, Yang S, Lin L, et al. Immune checkpoint inhibitors combined with angiogenic inhibitors in the treatment of locally advanced or metastatic lung adenocarcinoma patients. Cancer Immunology, Immunotherapy. 2023 Feb 1;72(2):449–59.
104. Figueiredo A, Almeida MA, Almodovar MT, Alves P, Araújo A, Araújo D, et al. Real-world data from the Portuguese Nivolumab Expanded Access Program (EAP) in previously treated Non Small Cell Lung Cancer (NSCLC). Pulmonology. 2020 Jan 1;26(1):10–7.
105. Fiorica F, Belluomini L, Stefanelli A, Santini A, Urbini B, Giorgi C, et al. Immune checkpoint inhibitor nivolumab and radiotherapy in pretreated lung cancer patients. American Journal of Clinical Oncology: Cancer Clinical Trials. 2018 Nov 1;41(11):1101–5.
106. Frost N, Kollmeier J, Misch D, Vollbrecht C, Grah C, Matthes B, et al. Pembrolizumab as First-Line Palliative Therapy in PD-L1 Overexpressing (≥ 50%) NSCLC: Real-world Results with Special Focus on PS ≥ 2, Brain Metastases, and Steroids. Clin Lung Cancer. 2021 Sep 1;22(5):411–22.
107. Fukui T, Okuma Y, Nakahara Y, Otani S, Igawa S, Katagiri M, et al. Activity of Nivolumab and Utility of Neutrophil-to-Lymphocyte Ratio as a Predictive Biomarker for Advanced Non–Small-Cell Lung Cancer: A Prospective Observational Study. Clin Lung Cancer. 2019 May 1;20(3):208-214.e2.
108. Geier M, Descourt R, Corre R, Léveiller G, Lamy R, Goarant É, et al. Real life second-line nivolumab in advanced non-small cell lung cancer: A French observational multicenter study of 259 patients (ABCT-IMMUNOBZH). Cancer Rep Rev. 2018;2(5).
109. Girard N, Smit HJM, Sibille A, et. al. PACIFIC-R real-world study: treatment duration and interim analysis of progression-free-survival in unresectable stage III NSCLC patients treated with durvalumab after chemoradiotherapy. 2021 p. Presented at: 2021 ESMO Congress; September 16-12.
110. Gong J, Gu J, Jiang L, Zhao D, Shao L, Chen X, et al. Clinical Study on the Efficacy of Bevacizumab in Combination with Pembrolizumab on Cellular Immune Function in the Treatment of Driver Gene-Negative Stage IV Lung Adenocarcinoma. J Oncol. 2022;2022.
111. Gu X, Shi Z, Shao L, Zhang Y, Zhang Y, Song Z, et al. Efficacy and safety of maintenance immune checkpoint inhibitors with or without pemetrexed in advanced non-squamous non-small cell lung cancer: a retrospective study. BMC Cancer. 2022 Dec 1;22(1).
112. Hong L, Lewis WE, Nilsson M, Patel S, Varghese S, Rivera MJ, et al. Limited Benefit from the Addition of Immunotherapy to Chemotherapy in TKI-Refractory EGFR-Mutant Lung Adenocarcinoma. Cancers (Basel). 2022 Jul 1;14(14).
113. Hsu JC, Lin JY, Hsu MY, Lin PC. Effectiveness and safety of immune checkpoint inhibitors: A retrospective study in Taiwan. PLoS One. 2018 Aug 1;13(8).
114. Hu R, Zhao Z, Shi Y, Shi M, Xia G, Yu S, et al. Immune checkpoint inhibitors combined with chemotherapy/ bevacizumab therapy for patients with advanced lung cancer and heavily treated with EGFR mutation: A retrospective analysis. J Thorac Dis. 2021 May 1;13(5):2959–67.
115. Hu J, Huang D, Wang Y, Li D, Yang X, Fu Y, et al. The efficacy of immune checkpoint inhibitors in advanced EGFR-Mutated non-small cell lung cancer after resistance to EGFR-TKIs: Real-World evidence from a multicenter retrospective study. Front Immunol. 2022 Sep 9;13.
116. Guberina M, Guberina N, Pöttgen C, Gauler T, Richlitzki C, Metzenmacher M, et al. Effectiveness of durvalumab consolidation in stage III non-small-cell lung cancer: focus on treatment selection and prognostic factors. Immunotherapy. 2022 Aug 1;14(12):927–44.
117. Guo X, Du H, Li J, Yang M, Xiong A, Zhang H, et al. Efficacy of ICIs on patients with oncogene-driven non-small cell lung cancer: a retrospective study. Cancer Drug Resistance. 2022;5(1):15–24.
118. He L, Chen X, Ding L, Zhang X. Clinical Efficacy of Antianlotinib Combined with Immune Checkpoint Inhibitors in the Treatment of Advanced Non-Small-Cell Lung Cancer and Its Effect on Serum VEGF, CEA, and SCC-Ag. J Oncol. 2022;2022.

119. Huang Y, Zhao JJ, Soon YY, Wong A, Aminkeng F, Ang Y, et al. Real-world experience of consolidation durvalumab after concurrent chemoradiotherapy in stage III non-small cell lung cancer. Thorac Cancer. 2022 Nov 1;13(22):3152–61.
120. Huang D, Cui P, Huang Z, Wu Z, Tao H, Zhang S, et al. Anti-PD-1/L1 plus anti-angiogenesis therapy as second-line or later treatment in advanced lung adenocarcinoma. J Cancer Res Clin Oncol. 2021 Mar 1;147(3):881–91.
121. Huang Z, Zhou C, Xiong Y, Yang F, Zeng F, Jiang W, et al. PD-1 inhibitor versus bevacizumab in combination with platinum-based chemotherapy for first-line treatment of advanced lung adenocarcinoma: A retrospective-real world study. Front Oncol. 2022 Nov 9;12.
122. Ikezawa Y, Mizugaki H, Morita R, Tateishi K, Yokoo K, Sumi T, et al. Current status of first-line treatment with pembrolizumab for non–small-cell lung cancer with high PD-L1 expression. Cancer Sci. 2022 Jun 1;113(6):2109–17.
123. Ismail RK, Schramel FMNH, van Dartel M, Pasmooij AMG, Cramer-van der Welle CM, Hilarius DL, et al. Individual patient data to allow a more elaborated comparison of trial results with real-world outcomes from first-line immunotherapy in NSCLC. BMC Med Res Methodol. 2023 Dec 1;23(1).
124. Ivanović M, Knez L, Herzog A, Kovačević M, Cufer T. Immunotherapy for Metastatic Non-Small Cell Lung Cancer: Real-World Data from an Academic Central and Eastern European Center. Oncologist. 2021 Dec 1;26(12):e2143–50.
125. Jung HA, Noh JM, Sun JM, Lee SH, Ahn JS, Ahn MJ, et al. Real world data of durvalumab consolidation after chemoradiotherapy in stage III non-small-cell lung cancer. Lung Cancer. 2020 Aug 1;146:23–9.
126. Kawachi H, Tamiya M, Taniguchi Y, Yokoyama T, Yokoe S, Oya Y, et al. Efficacy of Immune Checkpoint Inhibitor With or Without Chemotherapy for Nonsquamous NSCLC With Malignant Pleural Effusion: A Retrospective Multicenter Cohort Study. JTO Clin Res Rep. 2022 Jul 1;3(7).
127. Khan M, Zhao Z, Li X, Liao G. Anti-pd1 therapy plus whole-brain radiation therapy may prolong pfs in selected non–small cell lung cancer patients with brain metastases: A retrospective study. Int J Gen Med. 2021;14:8903–18.
128. Kim SH, Choi CM, Lee DH, Kim SW, Yoon S, Kim WS, et al. Clinical outcomes of nivolumab in patients with advanced non-small cell lung cancer in real-world practice, with an emphasis on hyper-progressive disease. J Cancer Res Clin Oncol. 2020 Nov 1;146(11):3025–36.
129. Kishi N, Matsuo Y, Shintani T, Ogura M, Mitsuyoshi T, Araki N, et al. Recurrence patterns and progression-free survival after chemoradiotherapy with or without consolidation durvalumab for stage III non-small cell lung cancer. J Radiat Res. 2023 Jan 20;64(1):142–53.
130. Kobayashi K, Nakachi I, Naoki K, Satomi R, Nakamura M, Inoue T, et al. Real-world Efficacy and Safety of Nivolumab for Advanced Non–Small-cell Lung Cancer: A Retrospective Multicenter Analysis. Clin Lung Cancer. 2018 May 1;19(3):e349–58.
131. Krefting F, Basara N, Schütte W, Späth-Schwalbe E, Alt J, Thiel S, et al. Clinical Experience of Immunotherapy Treatment: Efficacy and Toxicity Analysis of the Compassionate Use Program of Nivolumab in Patients with Advanced Squamous Cell Non-Small Cell Lung Cancer. Oncol Res Treat. 2019 May 1;42(5):243–54.
132. Kumar S, Joga S, Biswas B, Dabkara D, Prasad KT, Singh N, et al. Immune checkpoint inhibitors in advanced non–small cell lung cancer: A metacentric experience from India. Vol. 44, Current Problems in Cancer. Mosby Inc.; 2020.
133. Kuo CHS, Wang CC, Huang YC, Pavlidis S, Liu CY, Ko HW, et al. Comparison of a combination of chemotherapy and immune checkpoint inhibitors and immune checkpoint inhibitors alone for the treatment of advanced and metastatic non-small cell lung cancer. Thorac Cancer. 2019 May 1;10(5):1158–66.

134. Kurokawa K, Mitsuishi Y, Shimada N, Kawakami Y, Miura K, Miyawaki T, et al. Association between the efficacy and immune-related adverse events of pembrolizumab and chemotherapy in non-small cell lung cancer patients: a retrospective study. BMC Cancer. 2022 Dec 1;22(1).
135. Kwok WC, Cheong TF, Chiang KY, Ho JCM, Lam DCL, Ip MSM, et al. Clinical efficacy and safety of pemetrexed with or without either Bevacizumab or Pembrolizumab in patients with metastatic nonsquamous non–small cell carcinoma. Asia Pac J Clin Oncol. 2023 Feb 1;19(1):87–95.
136. Lang D, Huemer F, Rinnerthaler G, Horner A, Wass R, Brehm E, et al. Therapy Line and Associated Predictors of Response to PD-1/PD-L1-Inhibitor Monotherapy in Advanced Non-small-Cell Lung Cancer: A Retrospective Bi-centric Cohort Study. Target Oncol. 2019 Dec 1;14(6):707–17.
137. Lefebvre C, Martin E, Hendriks LEL, Veillon R, Puisset F, Mezquita L, et al. Immune checkpoint inhibitors versus second line chemotherapy for patients with lung cancer refractory to first line chemotherapy. Respir Med Res. 2020 Nov 1;78.
138. Li Y, Yang P, Zhou X, Yang X, Wu S. Programmed cell death 1 inhibitor plus chemotherapy vs. chemotherapy in advanced drive-gene-negative non-small-cell lung cancer patients: A real-world study. Front Surg. 2022 Sep 1;9.
139. Liao J, Liu C, Long Q, Wu X, Wang H, Yu H, et al. Direct Comparison Between the Addition of Pembrolizumab or Bevacizumab for Chemotherapy-Based First-Line Treatment of Advanced Non-Squamous Non-Small Cell Lung Cancer Lacking Driver Mutations. Front Oncol. 2021 Sep 29;11.
140. Liao G, Qian Y, Arooj S, Zhao Z, Yan M, Li Z, et al. Radiation Plus Anti-PD-1 Therapy for NSCLC Brain Metastases: A Retrospective Study. Front Oncol. 2021 Oct 21;11.
141. Liu Y, Gao Y, Wang Y, Zhao C, Zhang Z, Li B, et al. A single center analysis of first-line treatment in advanced KRAS mutant non-small cell lung cancer: real-world practice. BMC Cancer. 2022 Dec 1;22(1).
142. Long Y, Xiong Q, Song Q, Li Y, Li X, Qin B, et al. Immunotherapy plus chemotherapy showed superior clinical benefit to chemotherapy alone in advanced NSCLC patients after progression on osimertinib. Thorac Cancer. 2022 Feb 1;13(3):394–403.
143. Lu Z, Ye M, Sun T, Wu S, Lin Z, Zhang X, et al. Pembrolizumab for the better treatment of EGFR-mutant T790M-negative advanced lung adenocarcinoma patients than dual treatment of pemetrexed plus platinum after tyrosine kinase inhibitor treatment failure. Ann Palliat Med. 2022 Jun 1;11(6):2100–9.
144. Mao S, Zhou F, Liu Y, Yang S, Chen B, Xu J, et al. ICI plus chemotherapy prolonged survival over ICI alone in patients with previously treated advanced NSCLC. Cancer Immunology, Immunotherapy. 2022 Jan 1;71(1):219–28.
145. Matsumoto H, Kobayashi N, Somekawa K, Fukuda N, Kaneko A, Kamimaki C, et al. Pembrolizumab monotherapy versus pembrolizumab plus chemotherapy in patients with non-small-cell lung cancer: A multicenter retrospective trial. Thorac Cancer. 2022;13(2):228–35.
146. Miao K, Zhang X, Wang H, Si X, Ni J, Zhong W, et al. Real-World Data of Different Immune Checkpoint Inhibitors for Non-Small Cell Lung Cancer in China. Front Oncol. 2022 Mar 15;12.
147. Miura Y, Mouri A, Kaira K, Yamagguchi O, Shiono A, Hashimoto K, Nishihara F, Shinomiya S, Akagami T, Murayama, Y, Abe, T, Noda S, Kato S Kobayashi K KH. Thoracic Cancer - 2020 - Miura - Chemoradiotherapy followed by durvalumab in patients with unresectable advanced non-small.pdf. Thorac Cancer; 2020. p. 1280–7.

148. Molinier O, Besse B, Barlesi F, Audigier-Valette C, Friard S, Monnet I, et al. IFCT-1502 CLINIVO: real-world evidence of long-term survival with nivolumab in a nationwide cohort of patients with advanced non-small-cell lung cancer. ESMO Open. 2021 Feb [cited 2021 Dec 27];7(1):100353.
149. Morimoto K, Yamada T, Yokoi T, Kijima T, Goto Y, Nakao A, et al. Clinical impact of pembrolizumab combined with chemotherapy in elderly patients with advanced non-small-cell lung cancer. Lung Cancer. 2021 Nov 1;161:26–33.
150. Morimoto K, Uchino J, Yokoi T, Kijima T, Goto Y, Nakao A, et al. Early discontinuation of induction therapy in chemoimmunotherapy as an effective alternative to the standard regimen in patients with non-small cell lung cancer: a retrospective study. J Cancer Res Clin Oncol.2022 Sep;148(9):2437-2446.
151. Mouritzen MT, Junker KF, Carus A, Ladekarl M, Meldgaard P, Nielsen AWM, et al. Clinical features affecting efficacy of immune checkpoint inhibitors in pretreated patients with advanced NSCLC: a Danish nationwide real-world study. Acta Oncol (Madr). 2022;61(4):409–16.
152. Murteira R, Borges FC, Mendes GP, Ramos C, Ramos A, Soares P, et al. Real-world effectiveness of pembrolizumab in previously treated non-small cell lung cancer: A population-based cohort study. Pharmacoepidemiol Drug Saf. 2020 Oct 1;29(10):1295–302.
153. Nakamura Y, Miyazaki K, Aiko N, Misumi Y, Agemi Y, Taniguchi Y, et al. Efficacy of PD-1 inhibitors in older non-small cell lung cancer patients. Anticancer Res. 2020;40(2):923–8.
154. Offin M, Shaverdian N, Rimner A, Lobaugh S, Shepherd AF, Simone CB, et al. Clinical outcomes, local–regional control and the role for metastasis-directed therapies in stage III non-small cell lung cancers treated with chemoradiation and durvalumab. Radiotherapy and Oncology. 2020 Aug 1;149:205–11.
155. Park JH, You GL, Ahn MJ, Kim SW, Hong MH, Han JY, et al. Real-world outcomes of anti-PD1 antibodies in platinum-refractory, PD-L1-positive recurrent and/or metastatic non-small cell lung cancer, and its potential practical predictors: first report from Korean Cancer Study Group LU19-05. J Cancer Res Clin Oncol. 2021 Aug 1;147(8):2459–69.
156. Peng L, Guo J, Kong L, Huang Y, Tang N, Zhang J, et al. Efficacy of immunotherapy in KRAS-mutant advanced NSCLC: A real-world study in a Chinese population. Front Oncol. 2023 Jan 19;12.
157. Pons-Tostivint E, Hulo P, Guardiolle V, Bodot L, Rabeau A, Porte M, et al. Real-world multicentre cohort of first-line pembrolizumab alone or in combination with platinum-based chemotherapy in non-small cell lung cancer PD-L1 ≥ 50%. Cancer Immunology, Immunotherapy. 2023 Jun 1;72(6):1881–90.
158. Qiu L, Zhao X, Shi W, Sun S, Zhang G, Sun Q, et al. Real-world treatment efficacy of anti-programmed death-1 combined with anti-angiogenesis therapy in non-small cell lung cancer patients. Medicine (United States). 2020 Jun 12;99(24):E20545.
159. Qiao M, Zhou F, Hou L, Li X, Zhao C, Jiang T, et al. Efficacy of immune-checkpoint inhibitors in advanced non-small cell lung cancer patients with different metastases. Ann Transl Med. 2021 Jan;9(1):34–34.
160. Ratnayake G, Shanker M, Roberts K, Mason R, Hughes BGM, Lwin Z, et al. Prior or concurrent radiotherapy and nivolumab immunotherapy in non–small cell lung cancer. Asia Pac J Clin Oncol. 2020 Feb 1;16(1):56–62.
161. Rodríguez-Cid JR, Chards SCC, González-Espinoza IR, Garciá-Montes V, Garibay-Diáz JC, Hernández-Flores O, et al. A comparative study of immunotherapy as second-line treatment and beyond in patients with advanced non-small-cell lung carcinoma. Lung Cancer Manag. 2021 Sep 1;10(3).

162. Ruiz-Patiño A, Arrieta O, Cardona AF, Martín C, Raez LE, Zatarain-Barrón ZL, et al. Immunotherapy at any line of treatment improves survival in patients with advanced metastatic non-small cell lung cancer (NSCLC) compared with chemotherapy (Quijote-CLICaP). Thorac Cancer. 2020 Feb 1;11(2):353–61.
163. Saad A, Goldstein J, Appel S, Daher S, Urban D, Onn A, et al. Chemoradiation followed by adjuvant durvalumab in stage III non–small cell lung cancer: Real-world comparison of treatment outcomes to historical controls treated with chemoradiation alone. Thorac Cancer. 2022 Jun 1;13(12):1763–71.
164. Sabatier R, Nicolas E, Paciencia M, Jonville-Béra AP, Madroszyk A, Cecile M, et al. Nivolumab in routine practice for older patients with advanced or metastatic non-small cell lung cancer. J Geriatr Oncol. 2018 Sep 1;9(5):494–500.
165. Samaranayake C, McCaffrey E, Coucher J, Lehman M, Tao Mai G, Murphy M. Radiation therapy for augmenting the efficacy of immunotherapy in advanced non-small cell lung cancer: a case-controlled study. ERJ open res.2020 Mar 9;6(1);189-209
166. Samuel E, Lie G, Balasubramanian A, Hiong A, So Y, Voskoboynik M, et al. Impact of Radiotherapy on the Efficacy and Toxicity of anti-PD-1 Inhibitors in Metastatic NSCLC. Clin Lung Cancer. 2021 May 1;22(3):e425–30.
167. Sankar K, Bryant AK, Strohbehn GW, Zhao L, Elliott D, Moghanaki D, et al. Real World Outcomes versus Clinical Trial Results of Durvalumab Maintenance in Veterans with Stage III Non-Small Cell Lung Cancer. Cancers (Basel). 2022 Feb 1;14(3).
168. Schouten RD, Egberink L, Muller M, de Gooijer CJ, van Werkhoven E, van den Heuvel MM, et al. Nivolumab in pre-treated advanced non-small cell lung cancer: Long term follow up data from the Dutch expanded access program and routine clinical care. Transl Lung Cancer Res. 2020 Oct 1;9(5):1736–48.
169. Shen CI, Chao HS, Shiao TH, Chiang CL, Huang HC, Luo YH, et al. Comparison of the outcome between immunotherapy alone or in combination with chemotherapy in EGFR-mutant non-small cell lung cancer. Sci Rep. 2021 Dec 1;11(1).
170. Sheng J, Li H, Yu X, Yu S, Chen K, Pan G, et al. Efficacy of PD-1/PD-L1 inhibitors in patients with non-small cell lung cancer and brain metastases: A real-world retrospective study in China. Thorac Cancer. 2021 Nov 1;12(22):3019–31.
171. Shepard MJ, Xu Z, Donahue J, Eluvathingal Muttikkal TJ, Cordeiro D, Hansen L, et al. Stereotactic radiosurgery with and without checkpoint inhibition for patients with metastatic non–small cell lung cancer to the brain: a matched cohort study. J Neurosurg. 2020 Sep 1;133(3):685–92.
172. Shi Y, Ji M, Jiang Y, Yin R, Wang Z, Li H, et al. A cohort study of the efficacy and safety of immune checkpoint inhibitors plus anlotinib versus immune checkpoint inhibitors alone as the treatment of advanced non-small cell lung cancer in the real world. Transl Lung Cancer Res. 2022 Jun 1;11(6):1051–68.
173. Stenehjem DD, Lubinga SJ, Gupte-Singh K, Zhang Y, Le TK, Penrod JR, et al. Real-World Effectiveness of Nivolumab Monotherapy After Prior Systemic Therapy in Advanced Non–Small-Cell Lung Cancer in the United States. Clin Lung Cancer. 2021 Jan 1;22(1):e35–47.
174. Sun S, Liu C, Duan C, Yu S, Zhang Q, Xu N, et al. Efficacy and safety of immune checkpoint inhibitors in post-TKI NSCLC patients harboring EGFR mutations. J Cancer Res Clin Oncol. 2023 Jul 1;149(7):2937–49.
175. Takumida H, Horinouchi H, Masuda K, Shinno Y, Okuma Y, Yoshida T, et al. Comparison of time to failure of pembrolizumab plus chemotherapy versus pembrolizumab monotherapy: a consecutive analysis of patients having NSCLC with high PD-L1 expression. Cancer Immunology, Immunotherapy. 2022 Mar 1;71(3):737–46.

176. Taugner J, Käsmann L, Eze C, Tufman A, Reinmuth N, Duell T, et al. Durvalumab after chemoradiotherapy for PD-L1 expressing inoperable stage III NSCLC leads to significant improvement of local-regional control and overall survival in the real-world setting. Cancers (Basel). 2021 Apr 1;13(7).
177. Tian T, Yu M, Yu Y, Wang K, Tian P, Luo Z, et al. Immune checkpoint inhibitor (ICI)-based treatment beyond progression with prior immunotherapy in patients with stage IV non-small cell lung cancer: a retrospective study. Transl Lung Cancer Res. 2022 Jun 1;11(6):1027–37.
178. Tian T, Yu M, Li J, Jiang M, Ma D, Tang S, et al. Front-Line ICI-Based Combination Therapy Post-TKI Resistance May Improve Survival in NSCLC Patients With EGFR Mutation. Front Oncol. 2021 Nov 23;11.
179. Torasawa M, Yoshida T, Yagishita S, Shimoda Y, Shirasawa M, Matsumoto Y, et al. Nivolumab versus pembrolizumab in previously-treated advanced non-small cell lung cancer patients: A propensity-matched real-world analysis. Lung Cancer. 2022 May 1;167:49–57.
180. Trummer A, Bethge A, Dickgreber N, Dittrich I, Golpon H, Hoffknecht P, et al. NSCLC with uncommon EGFR mutations treated with atezolizumab plus bevacizumab and chemotherapy. Lung Cancer. 2022 Dec 1;174:141–5.
181. Tsai JS, Wei SH, Chen CW, Yang SC, Tseng YL, Su PL, et al. Pembrolizumab and Chemotherapy Combination Prolonged Progression-Free Survival in Patients with NSCLC with High PD-L1 Expression and Low Neutrophil-to-Lymphocyte Ratio. Pharmaceuticals. 2022 Nov 1;15(11).
182. Velcheti V, Chandwani S, Chen X, Pietanza MC, Piperdi B, Burke T. Outcomes of first-line pembrolizumab monotherapy for PD-L1-positive (TPS ≥50%) metastatic NSCLC at US oncology practices. Immunotherapy. 2019;11(18):1541–54.
183. Velcheti V, Hu X, Piperdi B, Burke T. Real-world outcomes of first-line pembrolizumab plus pemetrexed-carboplatin for metastatic nonsquamous NSCLC at US oncology practices. Sci Rep. 2021 Dec 1;11(1).9222
184. Vrankar M, Stanic K, Jelercic S, Ciric E, Vodusek AL, But-Hadzic J. Clinical outcomes in stage III non-small cell lung cancer patients treated with durvalumab after sequential or concurrent platinum-based chemoradiotherapy - Single institute experience. Radiol Oncol. 2021 Nov 19;55(4):482–90.
185. Vrdoljak E, Jakopović M, Geczi L, Bogos K, Bošković L, Magri C, et al. Real-World Safety and Efficacy of Nivolumab in Advanced Squamous and Nonsquamous Non-Small-Cell Lung Cancer: A Retrospective Cohort Study in Croatia, Hungary, and Malta. J Oncol. 2020;2020.
186. Wallrabenstein T, Mamot M, Daetwyler E, König D, Rothschild SI. Real-World Data of Combined Immunochemotherapy in Patients With Nonsquamous Advanced NSCLC. A Single-Center Retrospective Study. JTO Clin Res Rep. 2023 May 1;4(5).
187. Wang Y, Nie J, Dai L, Hu W, Zhang J, Chen X, et al. Evaluation of efficacy and toxicity of nivolumab combined with or without docetaxel in patients with advanced NSCLC. Cancer Immunology, Immunotherapy. 2022 Feb 1;71(2):267–76.
188. Wang M, Li J, Xu S, Li Y, Li J, Yu J, et al. Immunotherapy combined with chemotherapy improved clinical outcomes over bevacizumab combined with chemotherapy as first-line therapy in adenocarcinoma patients. Cancer Med. 2023 Mar 1;12(5):5352–63.
189. Wang X, Niu X, An N, Sun Y, Chen Z. Comparative Efficacy and Safety of Immunotherapy Alone and in Combination With Chemotherapy for Advanced Non-small Cell Lung Cancer. Front Oncol. 2021 Mar 18;11.
190. Wang W, Shao L, Xu Y, Song Z, Lou G, Zhang Y, et al. Efficacy and safety of anlotinib with and without EGFR-TKIs or immunotherapy in the treatment of elder patients with non-small-cell lung cancer: a retrospective study. BMC Pulm Med. 2022 Dec 1;22(1).
191. Wang P, Yin T, Zhao K, Yu J, Teng F. Efficacy of single-site radiotherapy plus PD-1 inhibitors vs PD-1 inhibitors for oligometastatic non-small cell lung cancer. J Cancer Res Clin Oncol. 2022 May 1;148(5):1253–61.

192. Wang Y, Zhang T, Wang J, Zhou Z, Liu W, Xiao Z, et al. Induction Immune Checkpoint Inhibitors and Chemotherapy Before Definitive Chemoradiation Therapy for Patients With Bulky Unresectable Stage III Non-Small Cell Lung Cancer. Int J Radiat Oncol Biol Phys. 2023 Jul 1;116(3):590–600.
193. Weis TM, Hough S, Reddy HG, Daignault-Newton S, Kalemkerian GP. Real-world comparison of immune checkpoint inhibitors in non-small cell lung cancer following platinum-based chemotherapy. Journal of Oncology Pharmacy Practice. 2020 Apr 1;26(3):564–71.
194. Wu K, Fu Y, Zeng D, Chen T, Wang C, Jiang J. A retrospective cohort study of sintilimab and pembrolizumab as first-line treatments for advanced non-small cell lung cancer. J Thorac Dis. 2022 Mar 1;14(3):679–88.
195. Xiong Q, Qin B, Xin L, Yang B, Song Q, Wang Y, et al. Real-World Efficacy and Safety of Anlotinib With and Without Immunotherapy in Advanced Non-Small Cell Lung Cancer. Front Oncol. 2021 Jul 29;11.
196. Xu B, Cheng H, Li K, Lv Y, Zeng X, Liu T, et al. Carboplatin and nab-paclitaxel chemotherapy with or without atezolizumab as front-line management for treatment-naïve metastatic nonsquamous non-small cell lung cancer with PD-L1 staining: a retrospective study. J Cancer Res Clin Oncol. 2022 Nov 1;148(11):3029–38.
197. Yang L, Hao X, Hu X, Wang Z, Yang K, Mi Y, et al. Superior efficacy of immunotherapy-based combinations over monotherapy for EGFR-mutant non-small cell lung cancer acquired resistance to EGFR-TKIs. Thorac Cancer. 2020 Dec 1;11(12):3501–9.
198. Yang Z, Chen Y, Wang Y, Hu M, Qian F, Zhang Y, et al. Pembrolizumab Plus Chemotherapy Versus Chemotherapy Monotherapy as a First-Line Treatment in Elderly Patients (≥75 Years Old) With Non-Small-Cell Lung Cancer. Front Immunol. 2022 Feb 14;13.
199. Li Y, Yang P, Zhou X, Yang X, Wu S. Programmed cell death 1 inhibitor plus chemotherapy vs. chemotherapy in advanced drive-gene-negative non-small-cell lung cancer patients: A real-world study. Front Surg. 2022 Sep 1;9.
200. Yu X, Chu X, Wu Y, Zhou J, Zhao J, Zhou F, et al. Favorable clinical outcomes of checkpoint inhibitorbased combinations after progression with immunotherapy in advanced non-small cell lung cancer. Cancer Drug Resistance. 2021;4(3):728–39.
201. Yu X, Li J, Ye L, Zhao J, Xie M, Zhou J, et al. Real-world outcomes of chemo-antiangiogenesis versus chemo-immunotherapy combinations in EGFR-mutant advanced non-small cell lung cancer patients after failure of EGFR-TKI therapy. Transl Lung Cancer Res. 2021 Sep 1;10(9):3782–92.
202. Zayas-Soriano M, Bonete-Sánchez M, Campillo-López J, Marcos-Ribes B, Hernández-Guio A, Aznar-Saliente MT. Clinical efficacy and safety of anti PD-1/PD-L1 antibodies as monotherapy in patients with non-small-cell lung cancer. Farmacia Hospitalaria. 2021 Jan 1;45(1):22–7.
203. Zhai X, Jing X, Li J, Tian Y, Xu S, Wang M, et al. Clinical Outcomes for PD-1 Inhibitor Plus Chemotherapy as Second-Line or Later Therapy Compared to PD-1/PD-L1 Inhibitor Alone in Advanced Non-small-cell Lung Cancer. Front Oncol. 2020 Sep 30;10.
204. Zhang F, Huang D, Li T, Zhang S, Wang J, Zhang Y, et al. Anti-PD-1 therapy plus chemotherapy and/or Bevacizumab as second line or later treatment for patients with advanced non-small cell lung cancer. J Cancer. 2020;11(3):741–9.
205. Zhang J, Wu D, Zhang Z, Long J, Tian G, Wang Y, et al. Pembrolizumab or Bevacizumab Plus Chemotherapy as First-Line Treatment of Advanced Nonsquamous Nonsmall Cell Lung Cancer: A Retrospective Cohort Study. Technol Cancer Res Treat. 2021;20.
206. Zhang T, Yang X, Zhao J, Xia L, Wang Q, Jin R, et al. The Application of Combined Immune Checkpoint Inhibitor Modalities in Previously Treated Non-Small Cell Lung Cancer Patients and the Associations Thereof With the Lung Immune Prognostic Index. Front Oncol. 2021 Jun 4;11.

207. Zhang X, Zeng L, Li Y, Xu Q, Yang H, Lizaso A, et al. Anlotinib combined with PD-1 blockade for the treatment of lung cancer: a real-world retrospective study in China. Cancer Immunology, Immunotherapy. 2021 Sep 1;70(9):2517–28.
208. Zhang W, Zhang C, Yang S, Chen Q, Wang C, Guo Q. Immune checkpoint inhibitors plus anlotinib versus anlotinib alone as third-line treatment in advanced non-small-cell lung cancer: A retrospective study. Future Oncology. 2021 Nov 1;17(31):4091–9.
209. Zhang Y, Zhu T, Wang Q, Wang J, Chen X. Original Article Effects of PD-1 inhibitor combined with anti-angiogenic drugs on efficacy and immune function of non-small cell lung cancer. Am J Transl Res. 2022.15;14(11):8225-8233.
210. Zhao Z, Hu R, Chen Y, Zhou G, Yu S, Feng J. Efficacy and Safety of PD-1 Immune Checkpoint Inhibitors in Locally Advanced and Advanced Non-Small-Cell Lung Cancer Patients with Chronic Infection. Oncol Res Treat. 2022 Jun 1;45(6):366–73.
211. Zhao X, Wu X, Yu H, Wang H, Sun S, Hu Z, et al. Prognostic value of hematologic parameters in advanced non-small cell lung cancer patients receiving anti-PD-1 inhibitors. Front Immunol. 2022 Oct 20;13.
212. Zheng MM, Tu HY, Yang JJ, Zhang XC, Zhou Q, Xu CR, et al. Clinical outcomes of non–small cell lung cancer patients with leptomeningeal metastases after immune checkpoint inhibitor treatments. Eur J Cancer. 2021 Jun 1;150:23–30.
213. Zhou S, Ren F, Meng X. Efficacy of immune checkpoint inhibitor therapy in EGFR mutation-positive patients with NSCLC and brain metastases who have failed EGFR-TKI therapy. Front Immunol. 2022 Sep 27;13.
214. Zhou ZC, Chen KY, Li N, Xie MY, Sheng JM, Fan Y, et al. Real-world utilization of PD-1/PD-L1 inhibitors with palliative radiotherapy in patients with metastatic non-small cell lung cancer. Thorac Cancer. 2022 Aug 1;13(16):2291–300.

# Supplementary

**Table S1.** Search strategy in Pubmed and Embase

| Database | Search strategy |
|---|---|
| PubMed | “Programmed Cell Death 1 Receptor”[Mesh] OR “programmed cell death 1 receptor” [tiab] OR “programmed cell death 1 receptors” [tiab] OR “Programmed Cell Death Protein 1 Inhibitor” [tiab] OR “Programmed Cell Death Protein 1 Inhibitors” [tiab] OR “pd-1 inhibitor” [tiab] OR “pd-1 inhibitors” [tiab] OR “PD-L1 inhibitor”[tiab] OR “PD 1 inhibitor” [tiab] OR “PD 1 inhibitors” [tiab] OR “PD-L1 inhibitors”[tiab] OR “Programmed Death-Ligand 1 Inhibitors” [tiab] OR “Programmed Death Ligand 1 Inhibitors” [tiab] OR “Programmed Death-Ligand 1 Inhibitor” [tiab] OR “Programmed Death Ligand 1 Inhibitor” [tiab] OR “CTLA-4 Antigen”[Mesh] OR “ctla-4 antigen” [tiab] OR “cytotoxic Tlymphocyte antigen 4” [tiab] OR “CTLA 4 inhibitors” [tiab] OR “CTLA 4 Inhibitors” [tiab] OR “CTLA- 4 Inhibitor” [tiab] OR “CTLA 4 Inhibitor” [tiab] OR “Cytotoxic T-Lymphocyte-Associated Protein 4 Inhibitors” [tiab] OR “Cytotoxic T Lymphocyte Associated Protein 4 Inhibitors” [tiab] OR “Cytotoxic T-Lymphocyte-Associated Protein 4 Inhibitor” [tiab] OR “Cytotoxic T Lymphocyte Associated Protein 4 Inhibitor” [tiab] OR “anti-pd-1” [tiab] OR “anti-pd-l1” [tiab] OR “anti-ctla-4” [tiab] OR “immune checkpoint inhibitor” [tiab] OR “immune checkpoint inhibitors” [tiab] OR “immune checkpoint blocker” [tiab] OR “immune checkpoint blockers” [tiab] OR “ICI” [tiab] OR “ici therapy” [tiab] OR “pembrolizumab” [tiab] OR “nivolumab”[tiab] OR “atezolizumab”[tiab] OR “durvalumab” [tiab] OR “ipilimumab” [tiab] OR “avelumab” [tiab] OR “tremelimumab” [tiab] OR “cemiplimab” [tiab] OR “Antibodies, Monoclonal, Humanized” [tiab]<br>AND<br>“Carcinoma, Non-Small-cell lung”[Mesh] OR NSCLC* [tiab] OR “non-small cell lung cancer*” [tiab] OR “non-small cell lung carcinoma*” [tiab] OR “Carcinoma, Non Small Cell Lung” [tiab] OR “Carcinomas, Non-Small-Cell Lung” [tiab] OR “Lung Carcinoma, Non-Small-Cell” [tiab] OR “Lung Carcinomas, Non-Small-Cell” [tiab] OR “Non-Small-Cell Lung Carcinomas” [tiab] OR “Nonsmall Cell Lung Cancer” [tiab] OR “Carcinoma, Non-Small Cell Lung” [tiab]<br>AND<br>“Progression-free survival”[Mesh] OR “Progression free survival” [tiab] OR “progression-free survival” [tiab] OR PFS [tiab] OR “progression-free” [tiab] OR “progression free” [tiab] OR “time to progression”[tiab] OR “Survival, Progression-Free” [tiab]” |
| Embase | “programmed death 1 receptor”/exp OR “cytotoxic T lymphocyte antigen 4”/exp OR ‘programmed cell death 1 receptor’:ab,ti OR ‘programmed cell death 1 receptors’:ab,ti OR ‘programmed cell death protein 1 inhibitor’:ab,ti OR ‘programmed cell death protein 1 inhibitors’:ab,ti OR ‘pd-l1 inhibitor’:ab,ti OR ‘pd 1 inhibitor’:ab,ti OR ‘pd 1 inhibitors’:ab,ti OR ‘pd-l1 inhibitors’:ab,ti OR ‘programmed death ligand 1 inhibitors’:ab,ti OR ‘programmed death ligand 1 inhibitor’:ab,ti OR ‘ctla 4 antigen’:ab,ti OR ‘cytotoxic t-lymphocyte antigen 4’:ab,ti OR ‘ctla 4 inhibitors’:ab,ti OR ‘ctla 4 inhibitor’:ab,ti OR ‘cytotoxic t lymphocyte associated protein 4 inhibitors’:ab,ti OR ‘cytotoxic t lymphocyte associated protein 4 inhibitor’:ab,ti OR ‘anti-pd-1’:ab,ti OR ‘anti-pd-l1’:ab,ti OR ‘anti-ctla-4’:ab,ti OR ‘immune checkpoint inhibitor’:ab,ti OR ‘immune checkpoint inhibitors’:ab,ti OR ‘immune checkpoint blocker’:ab,ti OR ‘immune checkpoint blockers’:ab,ti OR ‘ici’:ab,ti OR ‘ici therapy’:ab,ti OR ‘pembrolizumab’:ab,ti OR ‘nivolumab’:ab,ti OR ‘atezolizumab’:ab,ti OR ‘durvalumab’:ab,ti OR ‘ipilimumab’:ab,ti OR ‘avelumab’:ab,ti OR ‘tremelimumab’:ab,ti OR ‘cemiplimab’:ab,ti OR ‘antibodies monoclonal humanized’:ab,ti OR ‘immunotherapy’:ab,ti<br>AND<br>‘non small cell lung cancer’/exp OR ‘nsclc*’:ab,ti OR ‘non small cell lung cancer*’:ab,ti OR ‘non small cell lung carcinoma*’:ab,ti OR ‘carcinomas non small cell lung’:ab,ti OR ‘lung carcinoma non small cell’:ab,ti OR ‘lung carcinomas non small cell’:ab,ti OR ‘non-small-cell lung carcinomas’:ab,ti OR ‘nonsmall cell lung cancer’:ab,ti OR ‘carcinoma non small cell lung’:ab,ti<br>AND<br>‘progression free survival’/exp OR ‘progression free survival’:ab,ti OR ‘pfs’:ab,ti OR ‘progression-free’:ab,ti OR ‘time to progression’:ab,ti OR ‘survival progression free’:ab,ti<br>AND<br>[2011-2023]/py |

**Table S2.** Data extraction template

| | **Specification** |
|---|---|
| Name study | … |
| Number in search file | … |
| First author | … |
| Publication journal | … |
| Publication year | … |
| Journal impact factor | … |
| Initiated by pharmaceutical industry | Yes/no |
| Continent | America/Africa/Asia/Oceania/Multi-continent/Not reported |
| Type of study design | RCT/observational study |
| Type of site | Multicenter / Single center / other |
| | If multicenter, how many centers involved: … |
| Type of data | EHR/EHR data supplemented with other data/ Registries (treatment, disease or other, please specify)/ Claims data/ Clinical trial data/ Other, please specify /Not reported |
| Type of comparison | For RCTs;<br>- two treatment arms<br>- >two treatment arms<br><br>For observational studies<br>- one cohort versus trial estimate<br>- one cohort versus IPD trial<br>- two cohorts with contemporary treatments<br>- two cohorts: one with treatment of interest and one historical cohort<br>- > two cohorts (contemporary or historical) |
| Follow-up time (in months) | - For the total study population<br>- For the treatment of interest<br>- For the comparator treatment |
| Stage of disease | Stage IIIb/ Stage IIIc/ Stage IVa / Stage IVb / Other, please specify / Not reported |
| Type of histology | Squamous / Adeno / Large cell / Not other specified / Other, please specify / Not reported |
| Line of treatment | 1L / 2L/ higher lines/ Other, please specify / Not reported |
| Type of treatment of interest | Any PD-1/L1 or CTLA-4 inhibitor alone or in combination with other treatments. Specify all anticancer drugs used within one treatment regimen.<br>… |
| Type of comparator | Specify all comparators (any type of systemic treatment or trial estimates or IPD trial)<br>… |
| Cycle duration (in weeks) | - For the treatment of interest<br>…<br>- For the comparator treatment<br>… |
| Event definition | Disease progression and death / Disease progression alone / Other, please specify/ Not reported |

**Table S2.** Continued.

| | **Specification** |
|---|---|
| Start date for calculation PFS | Date of treatment start / Date of diagnoses / Date of randomization / Other, please specify / Not reported |
| End date for PFS calculation | Date of last follow-up / Date of last radiological scan / End of study period / Other, please specify / Not reported |
| Radiological and/or non-radiological criteria | Only radiological / Only non-radiological / Both / Not Reported |
| | If non-radiological criteria were used, please specify |
| Type of radiological criteria | RECIST v1.1 / iRECIST / irRECIST / imRECIST / modified criteria, please specify modifications / other criteria, please specify / Not reported |
| Type of radiological imaging | (PET-)CT / MRI / X-ray / Other, please specify / Not reported |
| Radiological imaging reviewer | Radiologist / Physician / Investigator / Other, please specify / Not reported |
| Response assessment schedule | … |

**Table S3.** Categorization of 'response assessment schedule'.

| **Category** | **Definition** | **Example** |
|---|---|---|
| Highly detailed | Response assessment schedules that are expressed in exact unite such as cycles, days, weeks, months, with no significant allowance for variation (+/- 7 days) | *'Response assessment occur every 6 weeks'* |
| Moderately detailed | Response assessment schedules that are expressed in exact unite such as cycles, days, weeks, months, with allowance for variation (> 7 days) | '*Response assessments occur every 2 to 3 months'* |
| Not detailed | Response assessments schedules without precise timeframes | '*Response assessments occur at the physician's discretion'* |
| Not reported | - | - |

**Table S4.** Completeness of reporting of PFS evaluation characteristics for each characteristics of observational studies.

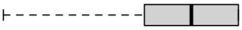
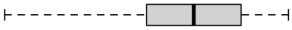
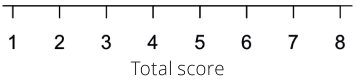
| **Study characteristics** | | **Observational studies (n = 144)** | **Completeness of reporting of PFS evaluation characteristics*** |
|---|---|---|---|
| | | **N (%)** | |
| Year of publication | 2015-2017 | 1 (1) | NA |
| | 2018-2020 | 43 (30) |  |
| | 2021-2023 | 100 (69) |  |
| Journal impact factor | < 4.4 | 72 (50) |  |
| | ≥ 4.4 | 72 (50) |  |
| Sponsored by the industry | Yes | 5 (3) |  |
| | No | 139 (97) |  |
| Continent | Asia | 90 (62) |  |
| | America | 12 (8) |  |
| | Europe | 36 (25) |  |
| | Australia | 3 (3) |  |
| | Multicontinent | 1(1) | NA |
| | Not reported | 2 (1) | NA |
| Site type | Multicenter | 53 (37) |  |
| | Single center | 89 (62) |  |
| | Not reported | 2 (1) | NA |
| Type of data | EHR | 74 (51) |  |
| | EHR combined with other data | 13 (9) |  |
| | Registry/database | 11 (8) |  |
| | Other | 2 (1) | NA |
| | Not reported | 44 (31) |  |
| Type of comparison | One cohort with trial (descriptive) | 34 (24) |  |
| | One cohort: Benchmark trial (statistical) | 4 (3) |  |
| | Two cohort: Contemporary | 82 (57) |  |
| | Two cohort: Historical | 8 (6) |  |
| | >Two cohort: Contemporary | 16 (11) |  |

**Table S4.** Continued.

| **Study characteristics** | | **Observational studies (n = 144)** | **Completeness of reporting of PFS evaluation characteristics*** |
|---|---|---|---|
| | | **N (%)** | |
| Follow-up time | Reported | 81 (56) |  |
| | Not reported | 63 (44) |  |
| Stage of disease | Stage III | 15 (10) |  |
| | Stage IV | 72 (52) |  |
| | Both | 40 (28) |  |
| | Not reported | 17 (12) |  |
| Histology | Non-squamous | 47 (33) |  |
| | Squamous | 2 (1) | NA |
| | Both | 91 (63) |  |
| | Not reported | 4 (3) |  |
| Line of treatment | 1L | 36 (25) |  |
| | 2L or later lines | 52 (36) |  |
| | Different lines | 42 (29) |  |
| | Not reported | 14 (10) |  |
| | | |  |

Abbreviations: EHR, electronic health records; NA, not applicable.
* For each study, the completeness of reporting for all PFS evaluation characteristics was assessed using a scoring system with a total score of 8. Each missing variable resulted in a reduction of 1 point from the total score.

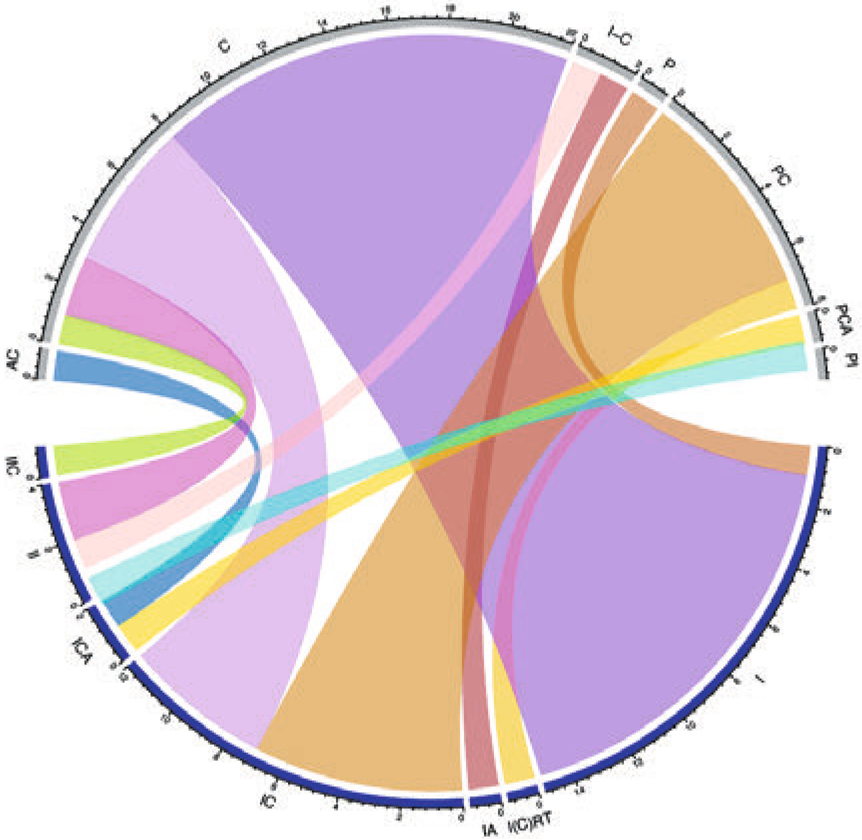


**Figure S1: Circle plot of treatment comparisons in clinical trial with two treatment arms (n = 36).** The segment borders are blue for the treatment of interest and grey for the comparator treatment.

Abbreviations: **AC**, anti-angiogenesis + chemotherapy; **C**, chemotherapy; **I**, immunotherapy; **IC**, immunotherapy + chemotherapy; **ICA**, immunotherapy + chemotherapy + anti-angiogenesis; **I(C)RT**, Immunotherapy + (chemo) radio therapy; **II**, Immunotherapy + immunotherapy; **IIC**, Immunotherapy + immunotherapy + chemotherapy; **P**, placebo; **PC**, placebo + chemotherapy; **PI**; Placebo + immunotherapy; **PCA** , placebo + chemotherapy + anti-angiogenesis.

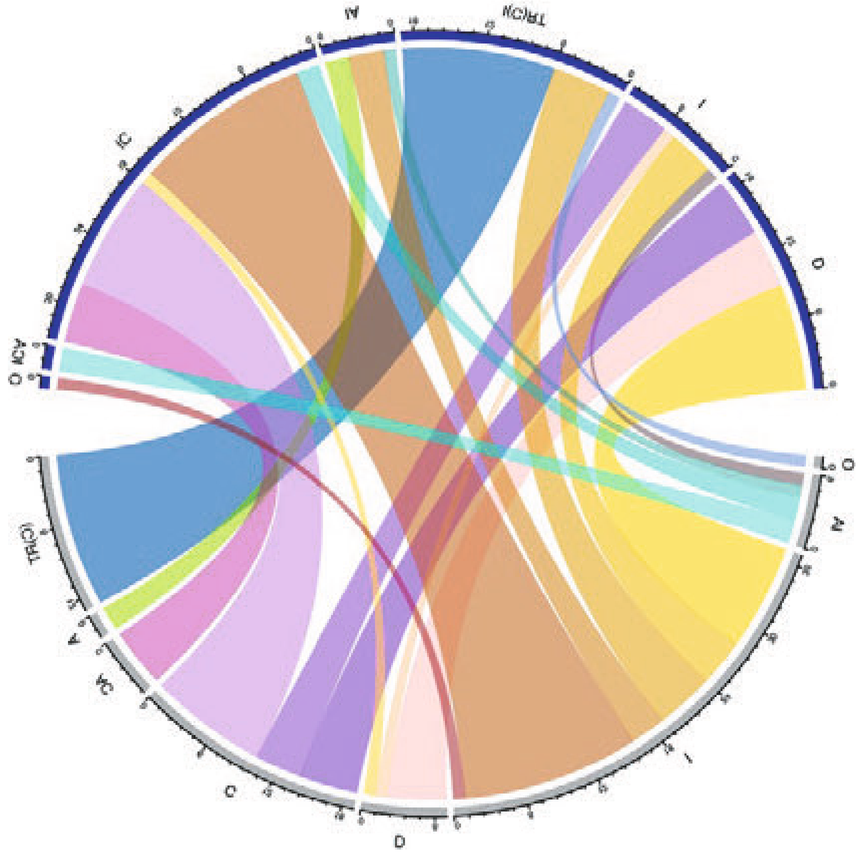


**Figure S2: Circle plot of treatment comparisons in observational studies with two treatment cohorts (n = 90).** The segment borders are blue for the treatment of interest and grey for the comparator treatment.

Abbreviations: **A**, anti-angiogenesis; **AC**, anti-angiogenesis + chemotherapy; **C**, chemotherapy; **D**, different treatments per arm; **ICA**, immunotherapy + chemotherapy + anti-angiogenesis; **I** , immunotherapy (monotherapy); **IA**, immunotherapy + anti-angiogenesis; **IC**, immunotherapy + chemotherapy; **I(C)RT**, immunotherapy + (chemo)radiotherapy; **O**, other.

# 3.2

# Development and portability of a text-mining algorithm for capturing disease progression in electronic health records of patients with stage IV non-small cell lung cancer treated with immunochemotherapy.

Marjon V Verschueren, Hanieh Abedian Kalkhoran, Maarten Deenen,
Ben EEM van den Borne, Juliëtte Visscher, Loes E Visser, Lourens T Bloem,
Bas JM Peters, Ewoudt MW van de Garde

*JCO Clinical Cancer Informatics. 2024 Oct 8:e2400053.*
*DOI: 10.1200/CCI.24.00053*

## Abstract

**Purpose**
The objective was to develop and evaluate the portability of a text-mining algorithm for prospectively capturing disease progression in electronic health record (EHR) data of patients with metastatic non-small cell lung cancer (mNSCLC) treated with immunochemotherapy.

**Methods**
This study used EHR data from mNSCLC patients receiving immunochemotherapy (between 01-10-2018 and 31-12-2022) in four Dutch hospitals. A text-mining algorithm for capturing disease progression was developed in hospitals 1 and 2 and then transferred to hospitals 3 and 4 to evaluate portability. Performance metrics were calculated by comparing its outcomes to manual chart review. Also, data were simulated to come available over time to assess performance in real-time applications. Median progression-free survival (mPFS) was calculated using the Kaplan-Meier method to compare text-mining to manual chart review.

**Results**
During development and portability, the text-mining algorithm performed well in capturing disease progression, with all performance scores >90%. When real-time performance was simulated, the performance scores in all four hospitals exceeded 90% from week 15 after the start of follow-up. Although the exact progression dates varied in 46 patients out of 157 patients with progressive disease, the number of patients labeled with progression too early (n = 24) and too late (n = 22) was well balanced with discrepancies ranging from -116 to 384 days. Nevertheless, the PFS curves constructed with text-mining and manual chart review were highly similar for each hospital.

**Conclusion**
In this study, an accurate text-mining algorithm for capturing disease progression in the EHR data of patients with mNSCLC was developed. The algorithm was portable across different hospitals, and the performance over time was good, making this an interesting approach for prospective follow-up of multicenter cohorts.

## Introduction

Real-world effectiveness studies of anti-cancer treatments commonly use progression-free survival (PFS) as an endpoint.[1,2] Most of these studies define PFS as the time from treatment initiation to the first documentation of disease progression or death, whichever comes first. [3] In clinical practice, disease progression is evaluated through a variety of assessments, such as imaging, physical examination, and pathological outcomes, which findings are primarily documented by the thoracic oncologist and radiologist in the unstructured text fields within the electronic health records (EHR). This multifactorial evaluation leads to a variety of terminology used to express disease progression (e.g., tumor growth, disease worsening, or clinical deterioration). The lack of structured data requires time-consuming manual chart review of multiple notes for capturing disease progression endpoints. According to Griffith et al., the median time required to extract progression events from the EHR for one patient was 18 minutes. [4] Therefore, manual review is not suitable for remote real-time detection of progression events in large cohorts under follow-up, as it would require repeated EHR reviews per patient over time, which is even more time-consuming. Remote multi-center real-time evaluations of time to disease progression in real-world settings, however, could bring relevant early warning signals for possible efficacy-effectiveness gaps of novel oncology medicines. However, alternative strategies for real-time follow-up studies are needed for efficient and reliable capturing progression endpoints in EHRs.

The challenges in collecting data from unstructured text fields have driven the exploration of alternative methods to identify and extract information. Text-mining and natural language processing (NLP) techniques have been used to transform free text into structured data, enabling more efficient detection of data items in EHRs. [5,6] Cheng et al. evaluated the performance of their custom-made NLP model for detecting tumor progression in brain tumor patients' neuroradiology reports, showing promising results with 87.3% sensitivity, 93.1% specificity, 84.9% positive predictive value, and 94.3% negative predictive value. [7] With the increasing success of custom-built NLP models, commercially available NLP is now beiding used more frequently to extract information from EHR text fields. Van Laar et al evaluated the performance of commercial NLP software in capturing patient characteristics and outcomes, including disease progression, in metastatic renal cell carcinoma patients. [8] The algorithm was able to capture disease progression, but the median PFS based on its output was 17% longer compared to manually retrieved dates of progression (8.9 vs. 7.6 months), indicating challenges in accurately detecting disease progression events. Furthermore, the study was conducted in a single-center setting, preventing conclusions about portability of their algorithm towards different EHR systems and clinical settings. Moreover, the algorithm's performance was evaluated using historical data, leaving uncertainty about the performance of prospective use for detecting progression events.

The objective of this study was to develop and evaluate the portability of a text-mining algorithm for prospectively capturing disease progression in EHRs of patients with mNSCLC and treated with immuno-chemotherapy in multicenter cohorts.

## Methods

### Study design, population and setting

The performance of a text-mining algorithm in capturing disease progression was assessed by comparing its results against manual chart review in a population of metastatic NSCLC (mNSCLC) patients treated with first-line pembrolizumab plus chemotherapy (immunochemotherapy). This population was chosen because the survival outcomes of mNSCLC are relatively short, and the treatment landscape is rapidly evolving, making it relevant to accelerate real-world evidence studies. Data was used from patients who received pembrolizumab plus chemotherapy between 01-10-2018 and 31-12-2022 in one of the following hospitals: Catharina Hospital Eindhoven (CZE; hospital 1), OLVG Hospital Amsterdam (OLVG; hospital 2), Leiden University Medical Centre (LUMC; hospital 3), and Haga Hospital Den Haag (HAGA; hospital 4). In hospital 1 and 2 the Santeon Farmadatabase and in hospital 3 and 4 the Clinical Data Collector (CTcue®) were used to identify eligible patients. For this study, the workflow was built up in three steps. In the first step ('the development') a text-mining algorithm was developed in hospital 1 and then transferred to hospital 2 for optimization. In the second step ('the portability') the algorithm was transferred to hospital 3 and 4 for external evaluation. Lastly, in step three ('the real-time simulation') data was made available at incremental time points, and the algorithm was evaluated at each point. There are no standards for reporting text-mining studies, so a modified version of the Standards for Reporting Diagnostic Accuracy Studies guidelines was used. [9]

### Ethics approval

In hospital 1 and 2, approval for patient data was obtained by the Santeon Institutional Review Board (SDB 2019-008). In hospital 3 and 4, approval was received by the local Medical Ethics Review Committees Leiden Den Haag Delft (P21-031). The need for informed consent was waived because of the retrospective nature of the study and most patients were deceased at time of conducting the study.

### Test methods

*Manual chart review (reference method)*

As the reference method, manual chart review was applied. Reviewer 1 (MV) conducted the review in hospitals 1 and 2, while reviewer 2 (HAK) performed the review in hospital 3 and 4. Both reviewers had expertise in the treatment of NSCLC patients. Manual chart review was conducted prior to text-mining, with a minimum interval of one week. In addition, the start date of pembrolizumab, date of death, and date of last clinic visit were retrieved with manual chart review.

*Text-mining (index test method)*

As an index test method, the CTcue® text-mining tool (IQVIA Patient Finder Solution-CTcue B.V., Amsterdam, The Netherlands)) was used. This Dutch language rule-based tool was employed to search and extract data from structured and unstructured EHR data, excluding PDF files. The tool employs NLP techniques to transform unstructured text into structured data. Data features that match the criteria of a query are subsequently presented in a graphical interface, presenting the identified text item with 150 characters before and after it by default. Subsequently, these identified data features require manual verification to ensure correctness. The underlying architecture of the tool is described in detail by Van Laar et al. (2020). [8]

## 1. Development

*Data source and retrieval*

For each patient, data were obtained from the EHR. Hospitals 2 employed EPIC software and hospitals 1, 3 and 4 employed Chipsoft Hix software as EHR systems. Information on disease progression was retrieved from all unstructured notes, such as medical letters, clinical or radiology notes. The period for retrieving information extended from the initial administration of pembrolizumab to the date of the last clinic visit for each patient.

*Variables of interest*

The variables of interest in this study were disease progression and the corresponding date of progression. To minimize potential bias due to inconsistent criteria for disease progression, a decision tree was developed to classify the presence and the date of disease progression (Figure S1). In patients with pseudoprogression, disease progression was only concluded when it was confirmed and reported in subsequent notes.

*Text-mining algorithm*

First, an initial text-mining algorithm was developed in hospital 1 by one researcher (MV), using clinical and radiological notes from ten randomly selected patients out of a total eligible population of 85 patients. Keywords related (e.g., disease progression, tumor progression, or tumor recurrence) and unrelated to progression (e.g., no progression, no signs of progression, or no recurrence) were identified through manual review. Additional words were formulated based on spelling errors, jargon, and literature terms. The list of included and excluded keywords is provided in Table S1a. The performance of the text-mining algorithm was evaluated in hospital 1 and afterwards transferred to hospital 2 for optimization. To ensure adaptability to different clinical settings, the same researcher (MV) randomly selected ten patients from the eligible population (n = 101) of hospital 2 and their clinical and radiological notes were also read entirely to identify new progression related words, see Table S1b, to update the text-mining algorithm. The performance of this updated algorithm was evaluated in hospital 2. Lastly, the final version of the algorithm was developed by analyzing the misclassifications hospital 1 and 2, which were used as input to update the list of (key)words (Table S1c).

### 2. Portability

The performance of the final version of the text-mining algorithm was evaluated on a total of 50 patients for each hospital (3 and 4) to gain insight into the portability. A random selection of patients was made from the eligible population using the Clinical Data Collector CTcue®, where patients are shuffled and pseudonymously assigned, ensuring a random sample. To assess the overall portability, the performance of both hospital 3 and 4 was evaluated after nine months of follow-up.

### 3. Real-time simulation

Through simulation of incremental data availability at weekly intervals, the algorithm's performance was assessed as if it were applied in real-time. The reference point for each patient was the start date of pembrolizumab (t=0), with EHR data made incrementally available with weekly intervals until the end of follow-up, i.e. either the date of last clinic visit or death. The performance of the text-mining algorithm for capturing disease progression was assessed for each interval (week 1; week 1 + 2; week 1 + 2 + 3; etc.).

## Analysis

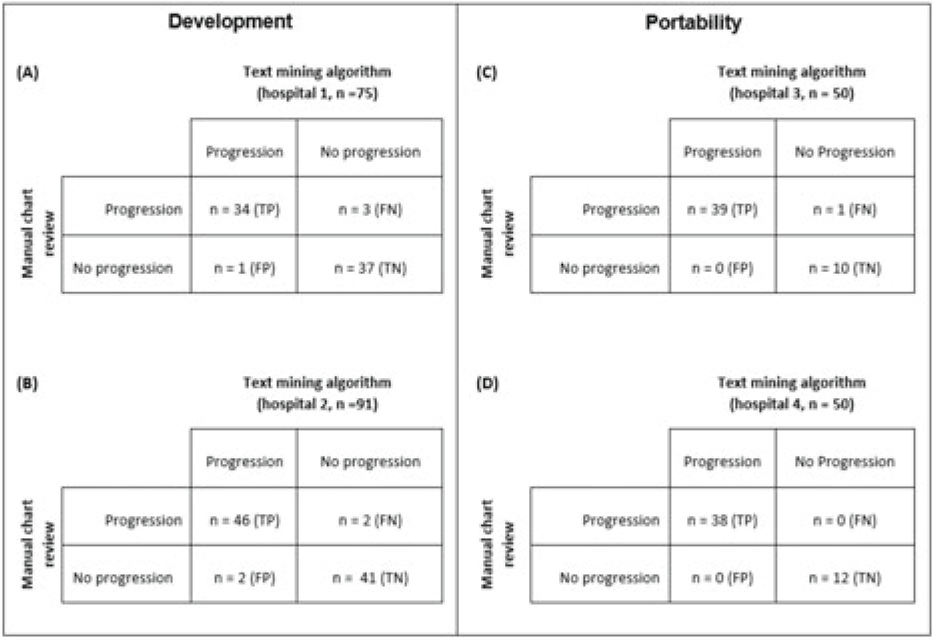
The text-mining results for disease progression were compared with manual chart review and classified as false positive (FP), false negative (FN), true positive (TP), and true negative (TN). The performance of the text-mining algorithm was evaluated using standard performance metrics (recall/sensitivity, precision/positive predictive value, F1-score, negative predictive value, and specificity); details are in Table S2 and S3. (11) Additionally, for TPs, a bar chart was constructed for the number of days between the date of disease progression extracted with text-mining and manual chart review. PFS was calculated from the start of treatment until progression or death, whichever occurred first. Patients without an event were censored on the date of their last clinic visit. The Kaplan-Meier method was used to visualize the PFS curve, estimate the median PFS (mPFS) and calculate the Hazard Ratio (HR) between the results captured using text-mining and manual chart review. All analyses were performed using R software version 1.4.2.

# Results

## Development

During development, the algorithm was applied on 75 patients in hospital 1 and 91 patients in hospital 2. In hospital 1, manual chart review identified 35 patients with and 40 patients without disease progression. Of these patients, the algorithm correctly captured 34 positives and 37 negatives, with one FN and three FPs (Table 1a). In hospital 2, manual chart review identified 48 patients with and 43 without disease progression. Of these patients, the updated algorithm correctly captured 46 positives and 41 negatives, with two FNs and two FPs (Table 1b). In both hospitals the text-mining algorithms scored high on all performance metrics, with all scores above 92% (Table 2).

**Table 1.** Confusion matrix for the performance of the text-mining queries during development **(A)** and **(B)** and portability **(C)** and **(D).**



**Development**

(A) Text mining algorithm (hospital 1, n =75)

| Manual chart review | Progression | No progression |
|---|---|---|
| Progression | n = 34 (TP) | n = 3 (FN) |
| No progression | n = 1 (FP) | n = 37 (TN) |

(B) Text mining algorithm (hospital 2, n =91)

| Manual chart review | Progression | No progression |
|---|---|---|
| Progression | n = 46 (TP) | n = 2 (FN) |
| No progression | n = 2 (FP) | n = 41 (TN) |

**Portability**

(C) Text mining algorithm (hospital 3, n = 50)

| Manual chart review | Progression | No Progression |
|---|---|---|
| Progression | n = 39 (TP) | n = 1 (FN) |
| No progression | n = 0 (FP) | n = 10 (TN) |

(D) Text mining algorithm (hospital 4, n = 50)

| Manual chart review | Progression | No Progression |
|---|---|---|
| Progression | n = 38 (TP) | n = 0 (FN) |
| No progression | n = 0 (FP) | n = 12 (TN) |

Abbreviations: FN, false negative; FP, false positive; TP, true positive; and TN, true negative.

**Table 2.** Overall performance during development and portability

| | Development | | Portability | |
|---|---|---|---|---|
| | **Hospital 1 (n = 75)** | **Hospital 2 (n = 91)** | **Hospital 3 (n = 50)** | **Hospital 4 (n = 50)** |
| Recall / sensitivity (%) | 97 | 96 | 98 | 100 |
| Precision / PPV (%) | 92 | 96 | 100 | 100 |
| F1-score (%) | 94 | 96 | 99 | 100 |
| NPV (%) | 97 | 95 | 91 | 100 |
| Specificity (%) | 93 | 95 | 100 | 100 |

Abbreviations: NPV, negative predictive value; PPV, positive predictive value.

For patients correctly captured with progression by text-mining the number of days between disease progression dates identified with manual chart review and text-mining are presented in Figures 1a and 1b. In hospital 1, 29% of patients (10/34) had date discrepancies (range: -89 to 25 days), with six patients being too late and four too early. In hospital 2, 33% of patients (15/46) had discrepancies (range: -116 to 384 days), with seven too late and eight too early. The reasons for date disagreements (for those exceeding 14 days) are outlined for each patient in Table S4. Additionally, in hospital 1 (Figure 2a) and hospital 2 (Figure 2b), no difference in the mPFS between manual chart review and text-mining was observed.

## Portability

The final version of the algorithm was transferred to hospital 3 (n = 50) and hospital 4 (n =50) to evaluate its portability, the results are presented in Table 1c and 1d respectively. In hospital 3, manual chart review identified 40 patients with and ten without progression. Of these patients, the algorithm correctly captured 39 positives and 10 negatives, with one FN identification. In hospital 4, manual chart review identified 38 patients with and 12 without progression. The algorithm correctly captured all these patients. The performance metrics in both hospitals were high ranging from 91% to 100% (Table 2).

In hospital 3, 23% (9/39) had date discrepancies (range: -6 to 77 days), with eight patients being too late and one being too early (Figure 1c). In hospital 4, 32% (12/38) had date discrepancies (range: - 11 to 17 days), with one too late and 11 too early (Figure 1d). The reasons for date disagreements (>14 days) are outlined for each patient in Table S4. In hospital 3 (Figure 2c), the mPFS estimated with manual chart review was shorter than the mPFS estimated with text-mining (5.3 versus 6.4 months), with no difference in the PFS curves. In hospital 4 (Figure 2d), there was no difference between the mPFS and PFS curves estimated with manual chart review and textmining (11 versus 11 months).

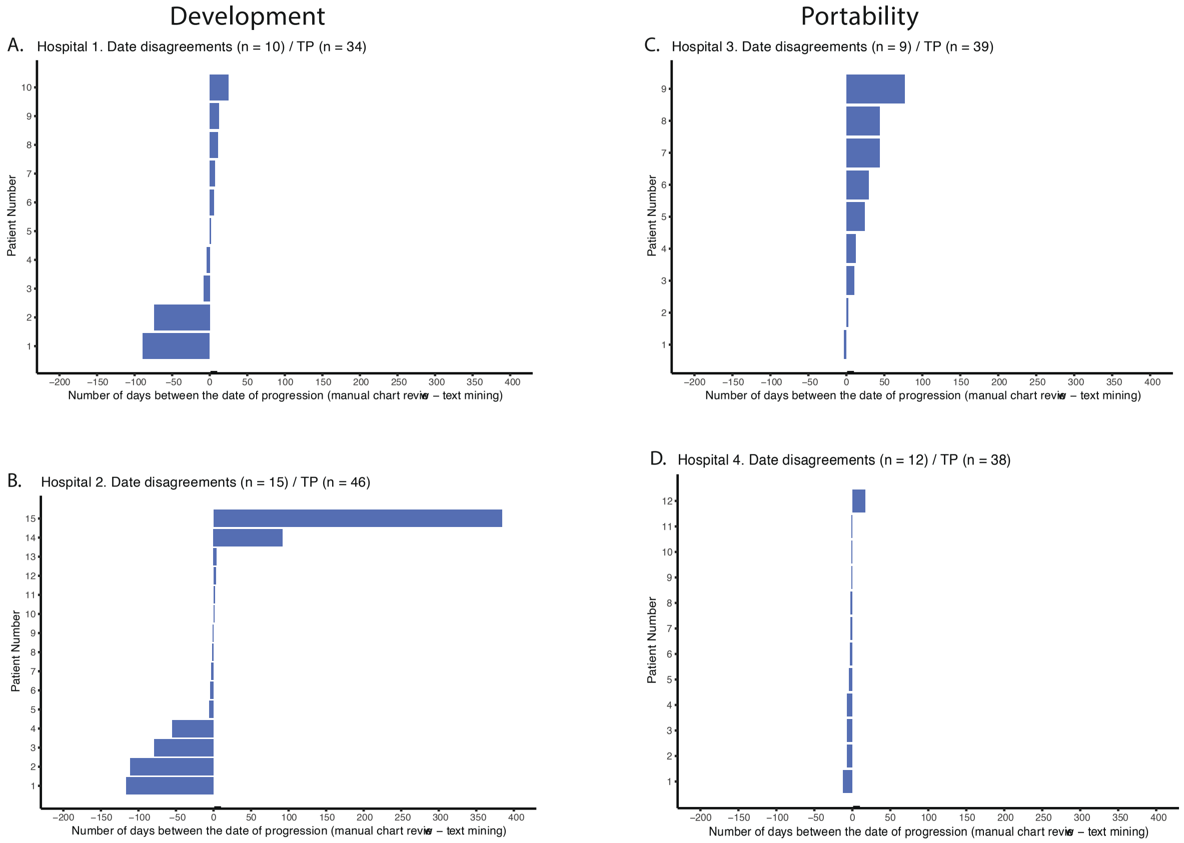


**Figure 1.** Horizontal bar chart presenting the number of days between the manually extracted date and text-mining date of progression for each patient correctly captured with progression during development in hospitals 1 **(A)** and 2 **(B),** and during portability in hospitals 3 **(C)** and 4 **(D).** Abbreviations: TPs, true positives.

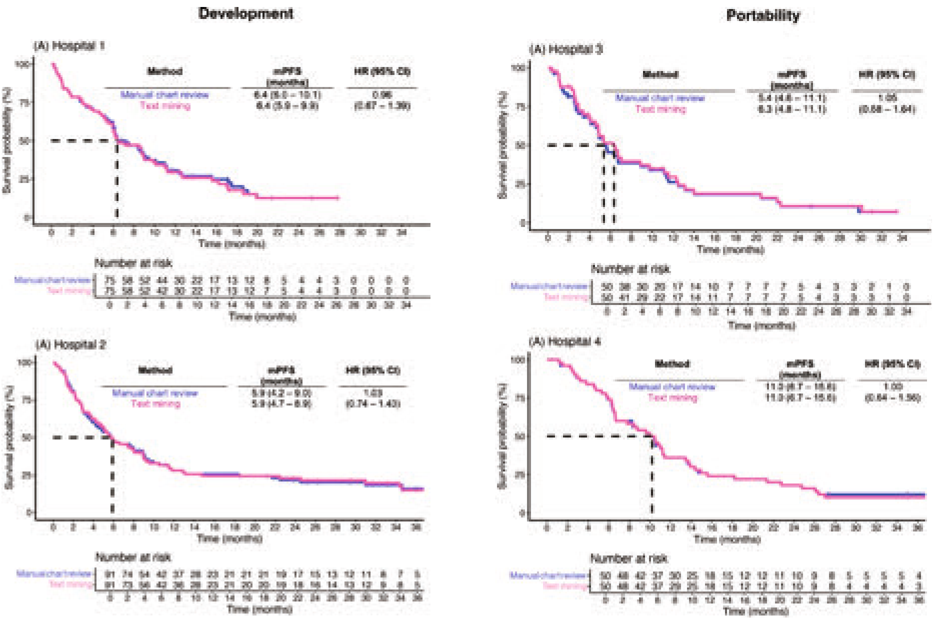


**Figure 2.** Kaplan-Meier curves of progression-free survival for data-extraction manual versus text-mining during development in hospital 1 (A) and 2 (B) and portability in hospital 3 (C) and 4 (D).

In summary, if the algorithm had been used to mNSCLC patients treated with first-line immunochemotherapy in hospitals 3 and 4 and evaluated atninemonths of follow-up, it would have only missed one patient with progression (FN) with no difference between the mPFS estimated with manual chart review and text-mining (6.7 versus 6.7 months).

### Real-time simulation

To simulate real-time application, data was provided incrementally every week (Table S4). In hospital 1, one FN occurred at week 12 and three FPs occurred at weeks 39, 71, and 74. In hospital 2, two FNs occurred at weeks 7 and 32, and two FPs occurred at week 15. Hospital 3 had only one FN at week 5, and hospital 4 had no incorrect identifications. Performance metrics for these simulations at each time interval are presented in Figure S2.

## Discussion

This study showed that our text-mining algorithm is feasible for capturing disease progression in the EHR of mNSCLC patients treated with immunochemotherapy. The algorithm performed well in all hospitals, with all performance scores higher than 91%. The algorithm also performed well over time when repeatedly applied to newly available EHR data at weekly intervals. Although the exact progression dates varied in approximately 30% of patients, the number of patients labeled with progression too early (n = 24) and too late (n = 22) was well balanced with discrepancies ranging from -116 to 384 days. Nevertheless, the PFS curves constructed with text mining and manual chart review were highly similar for each hospital. These outcomes indicate that this algorithm can be used to investigate the real-world effectiveness of cancer treatments.

This study is the first to develop and evaluate the performance of a text-mining algorithm for capturing disease progression in mNSCLC in multiple clinical settings. Previous studies on text-mining models for disease progression were either in different malignancies or single-center studies. For example, Cheng et al developed a text-mining algorithm for brain tumor patients with comparable performance scores (all scores > 87%) to our text-mining algorithm. [8] Ping et al captured recurrent disease in hepatocellular carcinoma patients with a comparable F1 score of 84%. [10] Recently, Lee et al. developed an algorithm called the ‘image-based-rule’, which uses text-mining to identify disease recurrence in radiological reports of patients with ovarian cancer. The performance of their algorithm was good, according to their reported numbers, an F1 score of 81% could be calculated. [11] Furthermore, our models’ overall performance was similar to the performance of Van Laar et al., who evaluated the same commercial NLP tool (CTcue®) for disease progression in renal cell carcinoma (F1 score 94.5%). [8] However, our algorithm was more accurate in estimating PFS curves with no difference from manual review, whereas the PFS curves estimated by Van Laar et al. were less identical, with a mPFS difference of 17% between text-mining and manual chart review. This difference may be due to our predefined decision tree for estimating progression, in contrast to their non-standardized approach. Additionally, they only included radiology notes, thus potentially missing relevant medical notes.

This study also evaluated the portability of the algorithm by testing its performance in multiple clinical settings. Overall, the algorithm performed well in different settings, with all performance scores above 90%. However, in 28% of all patients with progressive disease, discrepancies were observed between the calendar dates for progression. After analyzing the discrepancies, we observed that specific terms, such as ‘growth of lesions’ or ‘new nodular metastasis’, were used to indicate progression. This underscores that healthcare providers may use specific terminology to indicate progression. One solution to address this issue and improve generalizability is to manually review ten patients with and without progression to capture additional keywords before implementing the algorithm in a new setting and periodically after one year of deployment. Additionally, selecting patients with apparently short or long mPFS for manual review of their files can help users of the algorithm to identify potential false positives and negatives. Another explanation for the date discrepancies could be the lack of standardized response criteria, leading to varying interpretations of disease progression. This aligns with the findings of a previous study, demonstrating that relying only on standard response criteria to abstract information on cancer progression from EHRs is not possible, as these criteria were not used in 75% of the cases. [12] Discrete reporting of progression in standardized care pathways could overcome these disagreements. However, it is reassuring that no difference in the PFS curves was observed, indicating that the date disagreements were well balanced.

Another relevant finding is the good performance of the algorithm over time, indicating its potential for prospective follow-up of patients to capture disease progression. Currently, prospective patient follow-up involves repetitive manual chart review of newly available EHR data. However, this process is time-consuming. [4,13] Our algorithm could overcome this limitation by constructing an automated process for capturing progression events in multiple clinical settings parallel. The algorithm can run continuously, enabling real-time capture of progression, and facilitating real-time evaluation of PFS. This real-time evaluation allows for incremental and timely comparison with the PFS reported in clinical trials to assess potential efficacy-effectiveness gaps, which could provide an early signal to healthcare providers regarding of the relative effectiveness of specific treatments in clinical practice, enabling them to make informed decisions on treatment strategies. [14] However, sequentially evaluating this gap would require sophisticated statistical methodology to adjust for multiple testing to prevent false positive conclusions [15] and, the completeness and accuracy of mortality data should be validated for the calculation of PFS. [16] In the Netherlands, this validation is usually achieved by crosschecking it with the mortality data from the Personal Records Database. The study's strengths include its multicenter approach for developing and evaluating a text-mining algorithm, and assessing its performance over time. However, certain limitations should be addressed. First, the use of a commercial text-mining tool (CTcue®) lacks transparency in the used NLP methods, as the underlying algorithms and processing pipelines are not openly accessible. This limits the generalizability to Dutch-language hospitals using this commercial software. However, the identified keywords could also be used by non-commercial NLP models. Second, the data was not subject to a manual chart review in duplicate to ensure its quality. Nevertheless, a decision tree (Table S1) was used to minimize inconsistencies in the data, and in case of any ambiguities, a thoracic oncologist was consulted. Third, the mPFS observed in hospital 4 was around two times longer than the mPFS observed in the other hospitals, with no difference between manual chart review and text-mining (11 versus 11 months). As this study focused on the performance of our text-mining algorithm, a detailed analysis of factors contributing to this prolonged mPFS was not conducted. Factors that could contribute to PFS outcomes in mNSCLC patients treated with immunochemotherapy in clinical practice are described in more detail in our previous EE gap analysis. [17] Lastly, we only evaluated the performance of our algorithm in mNSCLC patients treated with pembrolizumab plus chemotherapy in the Netherlands. This limits the generalizability of our findings to other patient populations, treatment regimens, and geographical locations.

## Conclusion

In this study, an accurate text-mining algorithm for capturing disease progression in EHR data of patients with mNSCLC was developed. The algorithm was portable across different hospitals and the performance over time was good, making this an interesting approach for prospective follow-up of multicenter cohorts.

## Funding

This study was supported by Roche Nederland BV.

## References

1. Castelo-Branco L, Pellat A, Martins-Branco D, Valachis A, Derksen JWG, Suijkerbuijk KPM, et al. ESMO Guidance for Reporting Oncology real-World evidence (GROW). ESMO Real World Data and Digital Oncology. 2023 Nov;1:100003.
2. Arlett P, Kjaer J, Brioch K, Cooke E. Real-world evidence in EU medicines regulation: Enabling use and establishing value. Clin Pharmacol Ther. 2022;111(1):21–3.
3. Walia A, Tuia J, Prasad V. Progression-free survival, disease-free survival and other composite end points in oncology: improved reporting is needed. Nat Rev Clin Oncol. 2023 Dec 1;20(12):885–95.
4. Griffith SD, Miksad RA, Calkins G, You P, Lipitz NG, Bourla AB, et al. Characterizing the Feasibility and Performance of Real-World Tumor Progression End Points and Their Association With Overall Survival in a Large Advanced Non–Small-Cell Lung Cancer Data Set. JCO Clin Cancer Inform. 2019;(3):1–13.
5. Cai T, Giannopoulos AA, Yu S, Kelil T, Ripley B, Kumamaru KK, et al. Natural language processing technologies in radiology research and clinical applications. Radiographics. 2016;36(1):176–91.
6. Nadkarni PM, Ohno-Machado L, Chapman WW. Natural language processing: An introduction. Journal of the American Medical Informatics Association. 2011;18(5):544–51.
7. Cheng LTE, Zheng J, Savova GK, Erickson BJ. Discerning tumor status from unstructured MRI reports-completeness of information in existing reports and utility of automated natural language processing. J Digit Imaging. 2010;23(2):119–32.
8. van Laar SA, Gombert-Handoko KB, Guchelaar HJ, Zwaveling J. An Electronic Health Record Text Mining Tool to Collect Real-World Drug Treatment Outcomes: A Validation Study in Patients With Metastatic Renal Cell Carcinoma. Clin Pharmacol Ther. 2020;108(3):644–52.
9. Cohen JF, Korevaar DA, Altman DG, Bruns DE, Gatsonis CA, Hooft L, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: Explanation and elaboration. BMJ Open. 2016;6(11):1–17.
10. Ping XO, Tseng YJ, Chung Y, Wu YL, Hsu CW, Yang PM, et al. Information extraction for tracking liver cancer patients' statuses: From mixture of clinical narrative report types. Telemedicine and e-Health. 2013;19(9):704–10.
11. Lee S, Kim JH, Ha HI, Lim MC, Cho H. Development of an Automatic Rule-Based Algorithm for the Detection of Ovarian Cancer Recurrence From Electronic Health Records. JCO Clin Cancer Inform. 2024 Mar;(8).
12. Griffith SD, Tucker M, Bowser B, Calkins G, Chang C hsu, Guardino E, et al. Generating Real-World Tumor Burden Endpoints from Electronic Health Record Data: Comparison of RECIST, Radiology-Anchored, and Clinician-Anchored Approaches for Abstracting Real-World Progression in Non-Small Cell Lung Cancer. Adv Ther (Weinh). 2019;36(8):2122–36.
13. Griffith SD, Miksad RA, Calkins G, You P, Lipitz NG, Bourla AB, et al. Characterizing the Feasibility and Performance of Real-World Tumor Progression End Points and Their Association With Overall Survival in a Large Advanced Non–Small-Cell Lung Cancer Data Set. JCO Clin Cancer Inform. 2019;(3):1–13.
14. Nordon C, Karcher H, Groenwold RHH, Ankarfeldt MZ, Pichler F, Chevrou-Severac H, et al. The "efficacy-effectiveness gap": Historical background and current conceptualization. Value in Health. 2016;19(1):75–81.
15. Sjölander A, Vansteelandt S. Frequentist versus Bayesian approaches to multiple testing. Eur J Epidemiol. 2019;34(9):809–21.
16. Stewart M, Norden AD, Dreyer N, Joe Henk H, Abernethy AP, Chrischilles E, et al. An Exploratory Analysis of Real-World End Points for Assessing Outcomes Among Immunotherapy-Treated Patients With Advanced Non-Small-Cell Lung Cancer. JCO Clin Cancer Inform. 2019; 3:1-15.

17. Verschueren MV., Peters BJ, Bloem LT, Kruik VR, Uitvlugt EB, Bijsmans AR, et al. Pembrolizumab Plus Chemotherapy Per PD-L1 Stratum In Patients With Metastatic Non–Small Cell Lung Cancer: Real-World Effectiveness Versus Trial Efficacy. Clin Lung Cancer. 2024;25(2):119-127

# Supplementary

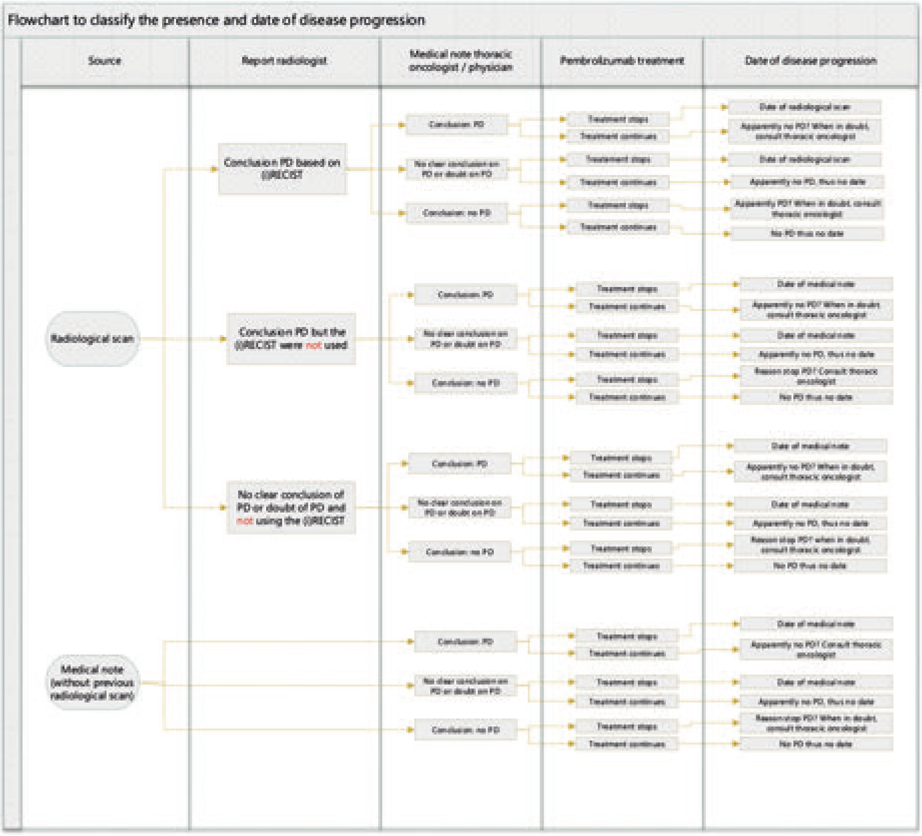


**Figure S1.** Flowchart to classify the presence and date of disease progression.

**Table S1.** Dutch (key)words included and excluded in the initial- (1a), updated- (1b) and final- (1c) text-mining algorithm. New (key)words identified during development are highlighted in grey.

| | Dutch (key)words related to 'progressive disease' used for the inclusion | Dutch (key)words not related to 'progressive disease' used for the exclusion | Time window for data extraction |
|---|---|---|---|
| **1.a. Initial text-mining algorithm** | progressie<br>progresie<br>progressive disease<br>progressieve<br>progressief<br>progresief<br>prgressie<br>prgrssie<br>prgressief<br>prgressieve<br>progressi<br>progressive<br>tumorprogressie<br>ziekteprogressie<br>ziekteprogresie<br>recidief<br>residief<br>residieve<br>recidieve<br>ziekterecidief<br>tumorrecidief | dd progressie<br>dd toch progressie<br>dd tumorprogressie<br>dd ziekteprogressie<br>suspectie op progressie<br>geen aanwijzingen voor progressie<br>geen aanwijzingen voor recidief<br>geen aanwijzingen voor tumorrecidief<br>geen aanwijzingen voor ziekteprogressie<br>dd tumorrecidief<br>dd ziekterecidief<br>progressieve pijn<br>progressieve weerstandsoefeningen<br>progressie pijn<br>progressie dyspneu<br>progressieve productieve hoest<br>geen progressie<br>geen progressieve<br>geen prgressie<br>geen progresie<br>geen recidief<br>geen recidieve<br>recidief: allogene<br>progressie 8xR-CHOP<br>geen aanwijzingen voor lokaal recidief<br>indien er sprake is van progressie<br>geen aanwijzingen voor lokaal recidief<br>dd bij algehele achteruitgang bij tumorprogressie<br>zorgen om progressie<br>verdacht voor recidief<br>progressie van vermoeidheid<br>recidief uitval<br>recidief klachten<br>progressie van leverfunctiestoornissen<br>progressieve leverfunctiestoornissen<br>recidief trombose<br>recidief pancreatitis | After the start date of pembrolizumabw |

**Table S1.** Continued.

| | Dutch (key)words related to 'progressive disease' used for the inclusion | Dutch (key)words not related to 'progressive disease' used for the exclusion | Time window for data extraction |
|---|---|---|---|
| **1.b. updated text-mining algorithm** | progressie<br>progresie<br>progressive disease<br>progressieve<br>progressief<br>progresief<br>prgressie<br>prgrssie<br>prgressief<br>prgressieve<br>progressi<br>progressive<br>tumorprogressie<br>ziekteprogressie<br>ziekteprogresie<br>recidief<br>residief<br>residieve<br>recidieve<br>ziekterecidief<br>tumorrecidief | dd progressie<br>dd toch progressie<br>dd tumorprogressie<br>dd ziekteprogressie<br>suspectie op progressie<br>geen aanwijzingen voor progressie<br>geen aanwijzingen voor recidief<br>geen aanwijzingen voor tumorrecidief<br>geen aanwijzingen voor ziekteprogressie<br>dd tumorrecidief<br>dd ziekterecidief<br>progressieve pijn<br>progressieve weerstandsoefeningen<br>progressie pijn<br>progressie dyspneu<br>progressieve productieve hoest<br>geen progressie<br>geen progressieve<br>geen prgressie<br>geen progresie<br>geen recidief<br>geen recidieve<br>recidief: allogene<br>progressie 8xR-CHOP<br>geen aanwijzingen voor lokaal recidief<br>indien er sprake is van progressie<br>geen aanwijzingen voor lokaal recidief<br>dd bij algehele achteruitgang bij tumorprogressie<br>zorgen om progressie<br>verdacht voor recidief<br>progressie van vermoeidheid<br>recidief uitval<br>recidief klachten<br>progressie van leverfunctiestoornissen<br>progressieve leverfunctiestoornissen<br>recidief trombose<br>recidief pancreatitis<br>recidief infarct<br>reciverende cerebrale infarcten<br>recidief CVA<br>recidief iCVA<br>recidief hereninfarc<br>recidief onder ticagrelor<br>R-top-progressie | After the start date of pembrolizumab |

**Table S1.** Continued.

| | Dutch (key)words related to 'progressive disease' used for the inclusion | Dutch (key)words not related to 'progressive disease' used for the exclusion | Time window for data extraction |
|---|---|---|---|
| **1.c. final text-mining query** | progressie<br>progresie<br>progressive disease<br>progressieve<br>progressief<br>progresief<br>prgressie<br>prgrssie<br>prgressief<br>prgressieve<br>progressi<br>progressive<br>tumorprogressie<br>ziekteprogressie<br>ziekteprogresie<br>recidief<br>residief<br>residieve<br>recidieve<br>ziekterecidief<br>tumorrecidief<br>hersenmetastasen nu weer aan het groeien | dd progressie<br>dd toch progressie<br>dd tumorprogressie<br>dd ziekteprogressie<br>suspectie op progressie<br>geen aanwijzingen voor progressie<br>geen aanwijzingen voor recidief<br>geen aanwijzingen voor tumorrecidief<br>geen aanwijzingen voor ziekteprogressie<br>dd tumorrecidief<br>dd ziekterecidief<br>progressieve pijn<br>progressieve weerstandsoefeningen<br>progressie pijn<br>progressie dyspneu<br>progressieve productieve hoest<br>geen progressie<br>geen progressieve<br>geen prgressie<br>geen progresie<br>geen recidief<br>geen recidieve<br>recidief: allogene<br>progressie 8xR-CHOP<br>geen aanwijzingen voor lokaal recidief<br>indien er sprake is van progressie<br>geen aanwijzingen voor lokaal recidief<br>dd bij algehele achteruitgang bij tumorprogressie<br>zorgen om progressie<br>verdacht voor recidief<br>progressie van vermoeidheid<br>recidief uitval<br>recidief klachten<br>progressie van leverfunctiestoornissen<br>progressieve leverfunctiestoornissen<br>recidief trombose<br>recidief pancreatitis<br>recidief infarct<br>reciverende cerebrale infarcten<br>recidief CVA<br>recidief iCVA<br>recidief hereninfarc<br>recidief onder ticagrelor<br>R-top-progressie | After the start date of pembrolizumab |

**Table S2.** Classification matrix

| Manual review | Text-mining | Classification |
|---|---|---|
| yes | yes | True positive (TP) |
| yes | no | False negative (FP) |
| no | no | True negative (TN) |
| no | yes | False positive (FP) |

**Table S3.** Performance metrics

$$\text{Recall/sensitivity} = \frac{true\ positives}{true\ positives + false\ negatives} \times 100\%$$

$$\text{Precision} — \text{Positive predictive value} \quad = \frac{true\ positives}{true\ positives + false\ positives} \times 100\%$$

$$\text{F1 score} = 2 \times \frac{recall \ \times \ precision}{recall \ + \ precision} \times 100\%$$

$$\text{Specificity} = \frac{true\ negatives}{true\ negatives + false\ positives}\ x\ 100\%$$

$$\text{Negative predictive value} = \frac{true\ negatives}{true\ negatives + false\ negatives} \times 100\%$$

**Table S4.** Reasons for progression date disagreement (when exceeding 14 days) between manual chart review and text mining algorithm for the patients with disease progression.

| | **Number of days between manual chart review and text mining** | **Reason text mining too early** | **Reason text mining too late** |
|---|---|---|---|
| **Hospital 1** | - 89 | **Missing information from subsequent notes**<br>The text mining algorithm found “slow progression” in a radiologist’s note, and this calendar date was identified as the date of progressive disease. However, subsequent clinical notes included information from the multidisciplinary meeting which concluded that the disease had not progressed. | |
| | - 74 | **Missing information from subsequent notes**<br>The text mining algorithm found “light progression” in a radiologist’s note and this calendar date was identified as the date of progression. However, the thoracic oncologist reported in the subsequent note that the treatment should be continued given that the total increase in lesions was less than 20%. | |
| | + 25 | | **Missing keyword(s)**<br>The text mining algorithm identified disease progression in a clinical note referring to “progression on imaging last month”. However, the algorithm failed to capture an earlier radiological note that mentioned “tumor growth” based on RECIST criteria, as these terms were not included as keywords. |
| **Hospital 2** | - 116 | **Missing contextual information**<br>The text mining algorithm captured the text “it looks like progression”; however it missed the contextual information provided earlier, which stated that the official results were not yet available. | |
| | - 111 | **Missing information from subsequent notes**<br>The text mining algorithm captured a radiological note with the text “disease progression”, but it turned out to be a misinterpretation of the radiological scan because the scan revealed immune-related pneumonitis instead. | |

**Table S4.** Continued.

| Number of days between manual chart review and text mining | Reason text mining too early | Reason text mining too late |
|---|---|---|
| - 79 | **Missing information from subsequent notes**<br>The text mining algorithm captured a clinical note stating, “slight progression of primary tumor.” However, in the subsequent notes, the care team discussed continuing the current first-line treatment (pembrolizumab was paused due to immune-related adverse events, and pemetrexed was continued as maintenance). | |
| - 55 | **Missing information from subsequent notes**<br>The text mining algorithm captured a clinical note with the text “Currently, radiological progression”, but it the subsequent notes, the presence of radiological progression is doubted by the multidisciplinary team. As a result, the decision was made to continue treatment until progressive disease was confirmed on the following scan. | |
| + 92 | | **Missing keyword(s)**<br>The text mining algorithm missed the clinical note with the text; ‘as a result of new nodular anomalies, stop immunotherapy, start best supportive care’. After 92 days, the algorithm captured the word disease progression in a clinical note, which was used to describe that progression had occurred in the past. |
| + 384 | | **Missing keyword(s)**<br>The text mining algorithm missed the clinical note describing disease progression with the text: “clinical and radiological deterioration, thus continue to second-line treatment”. The algorithm instead identified progression 389 days later based on a clinical note stating “disease progression” while on second-line treatment. |

**Table S4.** Continued.

| | Number of days between manual chart review and text mining | Reason text mining too early | Reason text mining too late |
|---|---|---|---|
| **Hospital 3** | + 24 | | **Missing keyword(s)**<br>The text mining algorithm failed to capture a radiological note with the text: “Significantly increased metabolic activity in the known right adrenal metastasis. This has also increased compared to the previous scan” |
| | + 29 | | **Missing keyword(s)**<br>The text mining algorithm failed to capture the radiological note with the text: “new nodular metastases” which was used to indicate disease progression. |
| | + 44 | | **Missing keyword(s)**<br>The text mining algorithm failed to capture a radiological note with the text: “growth of known nodules” which was used to indicate disease progression. |
| | + 44 | | **Missing keyword(s)**<br>The text mining algorithm failed to capture a radiological note with the text: “growth of metastases” which was used to indicate disease progression. |
| | + 77 | | **Missing keyword(s)**<br>The text algorithm failed to capture a radiological note with the text: “Multiple vertebral metastases, in retrospect, partly grown and partly new compared to previous examination” which was used to indicate disease progression. |
| **Hospital 4** | + 17 | | **Incomplete note**<br>The text mining algorithm could not capture disease progression because the thoracic oncologist had not completed the summary report. The report was eventually completed 17 days later including information about disease progression. |

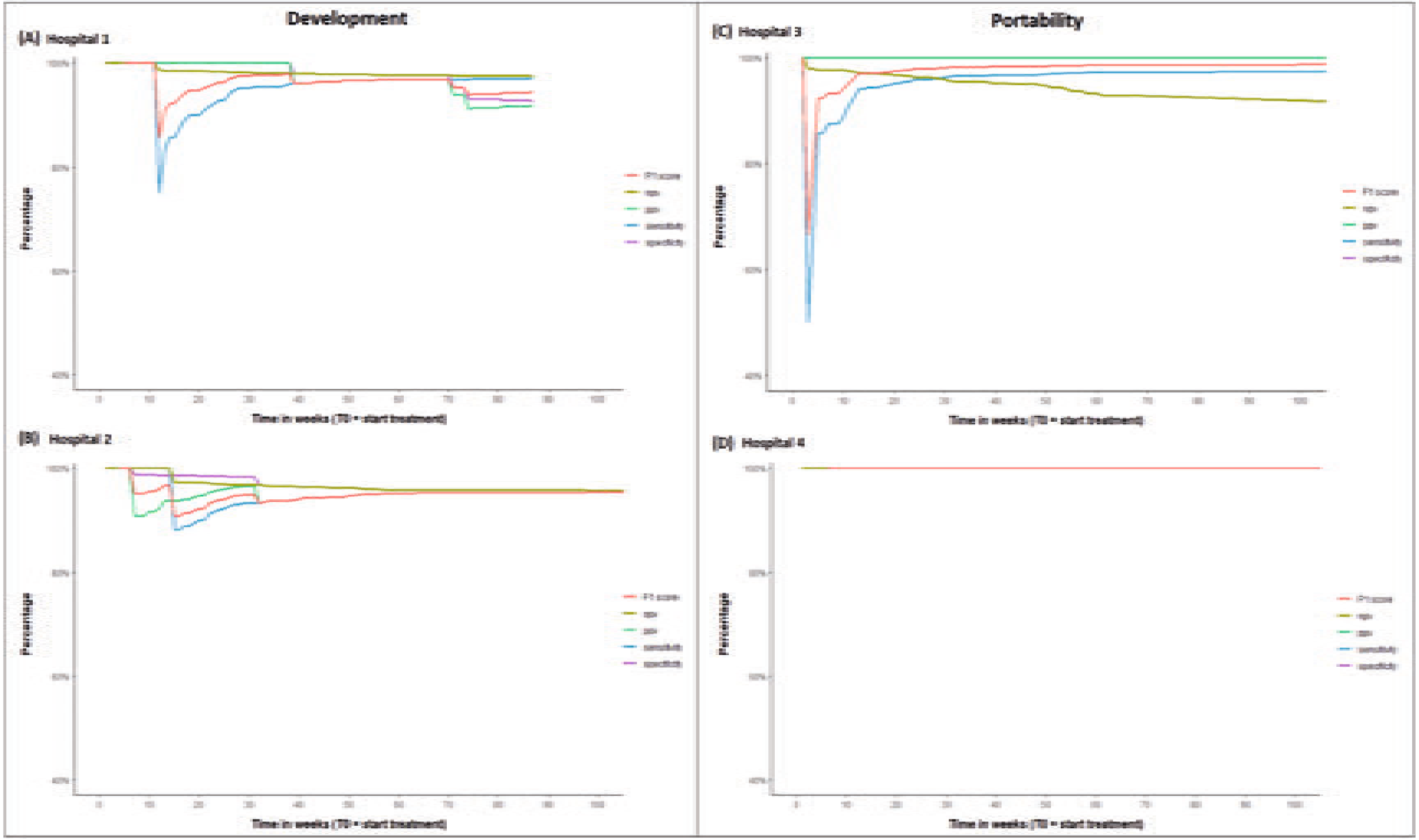


**Figure S2.** The F1 score, negative predictive value (NPV), positive predictive value (PPV), sensitivity and specificity over time for development in hospital 1 **(A)** and 2 **(B)** and portability in hospital 3 **(C)** and 4 **(D).**

# 3.3

# A Bayesian approach to compare accumulating time-to-event data from clinical practice with RCT data: a case study of overall survival in non-small cell lung cancer patients treated with immunotherapy.

Marjon V Verschueren, Daniel V Verschueren, Ewoudt MW van de Garde,
Lourens T Bloem.

*CPT Pharmacometrics & Systems Pharmacology. 2025 Dec; 14(12):2006-20013*
*DOI: 10.1002/psp4.70075.*

## Abstract

Survival outcomes observed in randomized controlled trials (RCTs) may not always be generalizable to clinical practice. Evaluating whether treatment outcomes in clinical practice are similar to those in RCTs shortly after a new medicine is introduced is important for making informed decisions. Therefore, we aimed to develop a Bayesian model that compares survival data from clinical practice that accumulates over time with static survival data from RCTs , thereby providing rapid and easily interpretable results that can inform clinical and policy-related decision-making. We developed a Bayesian survival model that sequentially updates estimates as new data become available. We designed the model to incorporate static RCT data with accumulating clinical practice data. We used sequential hypothesis testing with Bayes factors to assess the strength of the evidence for different hazard ratio (HR) thresholds (i.e., ranging from HR >1.0 to >2.0 and HR <0.5 to <1.0). We applied the model to two datasets comprising survival data from clinical practice and an RCT for lung cancer patients treated with pembrolizumab plus chemotherapy (dataset 1) and pembrolizumab monotherapy (dataset 2). For dataset 1, the posterior model checks showed a misfit between the model and the data after 15 months, potentially due to channeling bias. The model fit should be improved before reliable estimates can be obtained. For dataset 2, the model estimated precise HRs 10 months before the end of data accumulation. Sequential hypothesis testing with Bayes factors provided easily interpretable results, with very strong evidence for an HR > 1.0 and strong evidence for an HR > 1.2. In conclusion, provided the posterior check shows an acceptable model fit, our Bayesian survival model with sequential hypothesis testing using Bayes factors can provide rapid and easily interpretable results for decision-making.

## Introduction

Randomized controlled trials (RCTs) are considered the gold standard for determining the efficacy of new medicines.[1-3] However, outcomes observed in RCT populations may not always be generalizable to clinical practice. Patients treated in clinical practice may be older, less fit and have more comorbidities.[4] In addition, treatments and follow-up procedures are often less standardized and consistent in clinical practice than in RCTs. Therefore, it is important to evaluate whether treatment outcomes in clinical practice are similar to those in RCTs shortly after a new medicine is introduced.

Whether effectiveness observed in clinical practice and efficacy observed in RCT are similar has often been investigated especially for oncology medicines. Survival outcomes such as overall survival and progression-free survival are typically worse in clinical practice than in RCT populations. For example, the survival of patients with metastatic NSCLC treated with systemic treatments in clinical practice was shown to be nearly 25% shorter than the survival observed in RCTs.[4]

A difference between a medicine's efficacy (RCT) and effectiveness (clinical practice) is called an efficacy-effectiveness gap (EE gap).[5] The direction and size of an EE gap can vary depending on the type of medicine, setting and patient characteristics. In cases where an EE gap indicates a clinically relevant difference, it may warrant further investigation and adjustments to treatment strategies.

Time-to-event data, such as survival data, are typically modeled with Cox proportional hazards (PH) regression to estimate the hazard ratio (HR). In studies assessing whether there is an EE gap, the HR represents the hazard of an event occurring in the clinical practice cohort divided by that in the RCT cohort, while assuming that the HR is constant over time. An HR of 1 or close to 1 (e.g. an HR between 0.9 – 1.1) indicates similar outcomes for patients treated in clinical practice and those treated in an RCT, whereas an HR above this range indicates worse outcomes and an HR below this range indicates better outcomes for patients treated in clinical practice. However, estimating the appropriate sample size prior to a study is challenging due to variations in the type of medicine, setting and patient characteristics that may all influence whether treatment outcomes are similar or not. A too small sample size or too-short follow-up period may result in a lack of statistical power to detect clinically relevant differences between settings. Conversely, waiting until more patients have been treated or extending the follow-up period could incur unnecessary delays in investigating a potential EE gap.[6]

One potential solution is to compare accumulating clinical practice data repeatedly to RCT data. Bayesian survival models may offer a suitable approach for this type of repetitive evaluation because the Bayesian approach naturally allows sequential updates to model parameters and estimates. Unlike the frequentist Cox PH regression models, Bayesian models are designed to incorporate prior information explicitly, such as data from similar treatments and patient populations.[7-9] This Bayesian approach allows for the investigation of potential clinically relevant EE gaps as early as statistically justifiable, avoiding the a priori need for larger datasets or extended follow-up periods. Clinically relevant thresholds can be set to indicate that survival outcomes in clinical practice are worse or better than those reported in RCTs. Moreover, Bayesian models may allow a more intuitive interpretation of uncertainty than frequentist models, making it easier to assess and compare the confidence in estimates at different points in time.[10] This can support clinical and policy-related decision-making by indicating whether differences in survival outcomes between clinical practice and an RCT exist, enabling prompt adjustments to treatment strategies if needed.

Therefore, we aimed to develop a Bayesian survival model that compares survival data from clinical practice that accumulates over time with static survival data from RCTs, in order to investigate whether survival outcomes are similar or different thereby providing rapid and easily interpretable results that can inform for clinical and policy-related decision-making.

## Methods

In this section, we describe the key aspects of the development and application of our Bayesian survival model. Bayesian statistics fundamentally rely on Bayes' theorem, which includes three main elements: the observed information in the data, prior information, and posterior distribution. First, we explain the model development by detailing the observed information in the data, prior information, and data assimilation (step 1). Second, we describe the posterior distribution (step 2)**.** Third, we describe the model diagnostics we used to assess the performance of our model (step 3). Fourth, we explain how we used sequential hypothesis testing with Bayes Factors to evaluate potentially clinically relevant differences in survival outcomes (step 4), generating evidential strength in favor of one of the competing hypotheses. Fifth, we describe the prior sensitivity analyses used to assess the robustness of the estimates (step 5). Finally, we describe the two datasets that we used to apply the Bayesian survival model in more detail.

We implement the Bayesian survival model in Python using the PyMC package version 5.16.1 and use the Bayesian Analysis Reporting Guidelines (BARG) to ensure comprehensive and transparent reporting.[11] Full methodological details are available in the Supplementary Information, Files S1-S12.

## Step 1: Observed information in the data (likelihood), prior information and data assimilation

*Observed information in the data (likelihood)*

Our Bayesian survival model adopts key characteristics from Cox PH regression modeling. How these characteristics relate to Cox PH regression modeling is discussed in more detail in Supplementary File S1. The two primary model parameters of interest are the hazard rate and the baseline hazard, which are combined into a hazard function. Similar to Cox PH regression modeling, we assumed that the hazards are proportional and constant over time, and we accounted for right and non-informative censoring. While the baseline hazard does not need to be specified explicitly in Cox PH regression models, it must be defined in Bayesian models because of their fully parametric nature. Additionally, nuisance parameters for the baseline hazard should be explicitly defined in Bayesian survival modeling. These parameters and their distributions are important for constructing the likelihood function, which quantifies the probability of survival for the observed data. To model the baseline hazard, we used a Weibull distribution with two nuisance parameters $b$ (scale) and $k$ (shape) to model the baseline hazard.[12-14] Additionally, we incorporated one covariate in our model, representing the setting (clinical practice or RCT). Detailed mathematical information can be found in Supplementary File S2 and a synthetic validation of our model using simulated data is described in Supplementary File S3.

*Prior information*

Next, we defined weakly informed priors for these three model parameters, to provide some guidance without being overly restrictive. We used a normal distribution for the log hazard rate with $\mu = 0$ and a $\sigma = 0.5$, with a lognormal distribution for the scale parameter $b$ with $\mu = -3.0$ and $\sigma = 1.0$, and a uniform distribution for the shape parameter $k$ (with $k < 1$ indicating the rate decreasing over time and $k > 1$) indicating the rate increasing over time), with a range from 0.9 to 1.1. These prior distributions are plotted in Figure S1.

*Data assimilation*

The RCT and clinical practice cohorts and associated datasets are described in more detail below. We designed the Bayesian survival model to incorporate the data from the RCT cohort in a fixed manner while accumulating data from the clinical practice cohort over time as new patients are chronologically enrolled and their follow-up periods prolonged based on their treatment start and end dates. We updated the model at biweekly intervals, incorporating both newly enrolled patients and longer follow-up of already enrolled patients. Patients who were still alive by the end of follow-up were censored at the date of their last clinic visit, with the censoring date updated as the follow-up period was prolonged.

### Step 2: Posterior distribution

We used Monte Carlo Markov chain (MCMC) simulations to estimate the posterior probability distributions (Supplementary File S2 and S4). To present the results in an easily interpretable manner, we estimated the HRs over time and their corresponding 50%, 75%, 90%, 95%, and 97.5% credible intervals, which were derived from the posterior distribution. The Bayesian credible intervals provide an easily interpretable result, as they express the probability of a parameter falling within a specific range: a 95% credible interval reflects a 95% probability that the true value lies within that interval based on the accumulated data at a specific point in time.

### Step 3: Model diagnostics

Kaplan-Meier curves were constructed to visualize the data and to check whether the HRs are proportional and constant over time. Prior predictive checks were performed to evaluate the suitability of the chosen priors, ensuring alignment with data distributions before model fitting (Supplementary File S5). After applying the model, posterior predictive checks were used to assess the model fit, comparing model-generated data with the observed data (Supplementary File S6). Good model fits are necessary to obtain reliable estimates.

### Step 4: Bayesian hypothesis testing

To determine whether any observed differences in survival outcomes between patients treated in clinical practice and those in the RCT were clinically relevant, evidence for clinically relevant HR thresholds was calculated at each model update. HR thresholds from >1.0 to >2.0 were used to indicate worse survival outcomes (i.e., higher risk of death) in clinical practice compared to the RCT, while thresholds from <0.50 to <1.00 were used to indicate better survival outcomes (i.e., lower risk of death).

However, repeated evaluations can lead to multiple testing issues, increasing the likelihood of false positive conclusions. To mitigate this issue, we used the Bayes factor that quantifies the strength of the evidence provided by the data for the null hypothesis $H_0$ compared to the evidence for the alternative hypothesis $H_1$. A Bayes factor of 1 signifies inconclusive evidence for $H_1$, a Bayes factor between 1 and 3 indicates anecdotal evidence for $H_1$, a Bayes factor between 3 and 10 suggests moderate evidence for $H_1$, a Bayes factor between 10 and 30 indicates strong evidence for $H_1$, and a Bayes factor greater than 30 represents very strong evidence for $H_1$.[15] The Bayes factor is updated as new data are available, providing insight into which hypothesis ($H_0$ or $H_1$) is better supported by the evidence for every updated model.[15] To provide interpretable results at each model update, the Bayes factors with the different levels of evidence were plotted for all competing hypotheses testing different HR thresholds. For hypotheses that consistently reached very strong evidence in consecutive model updates, the graph was limited to this level. Details on Bayes factor hypothesis testing are reported in Supplementary File S7.

### Step 5: Prior sensitivity analysis

To evaluate the robustness of the estimates, sensitivity analyses with different priors and a different baseline hazard were conducted (Supplementary File S8 and S9). For the baseline hazard parameterization, a Gompertz distribution (instead of a Weibull distribution) was used.

*Application of the Bayesian survival model to two datasets*

To evaluate the performance of the Bayesian survival model in multiple settings, the model was applied to two datasets of patients with metastatic NSCLC (mNSCLC) treated with either pembrolizumab plus chemotherapy (dataset 1) or pembrolizumab monotherapy (dataset 2). These datasets were readily available and therefore suitable to evaluate the performance of the model.

#### Design and setting of dataset 1: Pembrolizumab plus chemotherapy

The first dataset contains survival data from two cohorts of patients with mNSCLC who were treated with first-line pembrolizumab plus chemotherapy: one derived from the phase III KEYNOTE-189 RCT (n = 410)[16], with survival data reconstructed with the algorithm of Guyot *et al.* [17], and one clinical practice cohort (n = 512) derived from electronic health records from the same seven non-university Dutch teaching hospitals of the Santeon Group. The RCT included patients treated and followed-up between February 2016 and March 2017, with a median follow-up of 10.5 months. The clinical practice cohort included patients treated and followed up between January 2019 and December 2021, with a median follow-up of 15.5 months. Previous research from Verschueren *et al.* (2024) demonstrated that the survival outcomes for patients treated in the clinical practice cohort were worse than the survival outcomes for patients treated in the RCT, as assessed with Cox PH regression modeling (HR 1.50; 95% confidence interval 1.26 – 1.79).[18] The distribution of patient and disease characteristics over time is described in Supplementary File S10 Table S3. More details regarding the clinical practice cohort and data analysis can be found in their publication.

#### Design and setting of dataset 2: Pembrolizumab monotherapy

The second dataset contains survival data from two cohorts of patients with mNSCLC who were treated with first-line pembrolizumab: one derived from the phase III KEYNOTE-024 RCT (n = 154)[19] and one clinical practice cohort (n = 83) derived from electronic health records from seven non-university Dutch teaching hospitals (Santeon Group). We used the algorithm developed by Guyot *et al.* to reconstruct survival data from the Kaplan-Meier curves of the RCT.[17] The RCT included patients treated and followed-up between September 2014 and October 2015, with a median follow-up of 11.2 months. The clinical practice cohort included patients treated and followed up between January 2015 and January 2019, with a median follow-up of 15.5 months. Previous research from Cramer *et al.* (2021) demonstrated that the survival outcomes for patients treated in the clinical practice cohort were worse than the survival outcomes for patients treated in the RCT, as assessed with Cox PH regression modeling (HR 1.55; 95% confidence interval 1.07 – 2.25).[20] The distribution of patient and disease characteristics over time is described in Supplementary File S10 Table S4. More details regarding the clinical practice cohort and data analysis can be found in their publication.

## Results

### Dataset 1: Survival outcomes associated with pembrolizumab plus chemotherapy

*Kaplan-Meier curves*

The Kaplan-Meier curves of overall survival outcomes associated with pembrolizumab plus chemotherapy for the RCT cohort and the complete clinical practice dataset (i.e., comprising of all patients and complete follow-up) are presented in Figure 1.

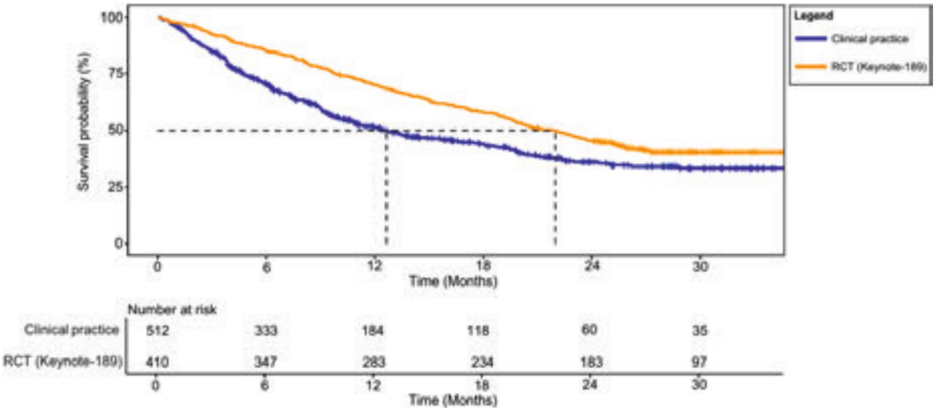


**Figure 1.** Kaplan-Meier curves of overall survival outcomes for patients treated with pembrolizumab plus chemotherapy in clinical practice versus the KEYNOTE-189 RCT.

*Model diagnostics*

The MCMC procedure demonstrated good convergence and resolution, as detailed in Table S1 and Figure S3. Figure S5 shows that the predictive prior checks demonstrated good alignment between the prior and the observed data, indicating that the chosen priors are appropriate. Figure S7a-j provide a visualization of the model fit at subsequent time intervals, through posterior predictive checks. Initially there was a good overlap between the data and the posterior, indicating a good model fit. However, after 15 weeks of data accumulation, there was a consistent mismatch between the data and the posterior model fit, suggesting that the model may be unable to capture certain complexities in the data over time. This indicates that the model should be improved (for example, through including additional covariates) before reliable estimates can be obtained and any meaningful conclusions can be drawn

*Posterior distribution*

Figure 2a shows the HR estimates for each consecutive model with credible intervals from 50% to 97.5%. Initially, in the first 15 months of data accumulation, the HR estimates fluctuated but showed an overall upward trend towards 2.20, indicating worse survival outcomes for patients treated in clinical practice compared to those treated in the RCT. As more data accumulated (i.e., more patients and longer follow-up), HR estimates decreased and reached a plateau at around 1.5 after approximately 35 months. When all data were included, the HR was 1.54 (95% credible interval 1.33 – 1.78) (Supplementary File S11, Table S5).

*Bayesian hypothesis testing*

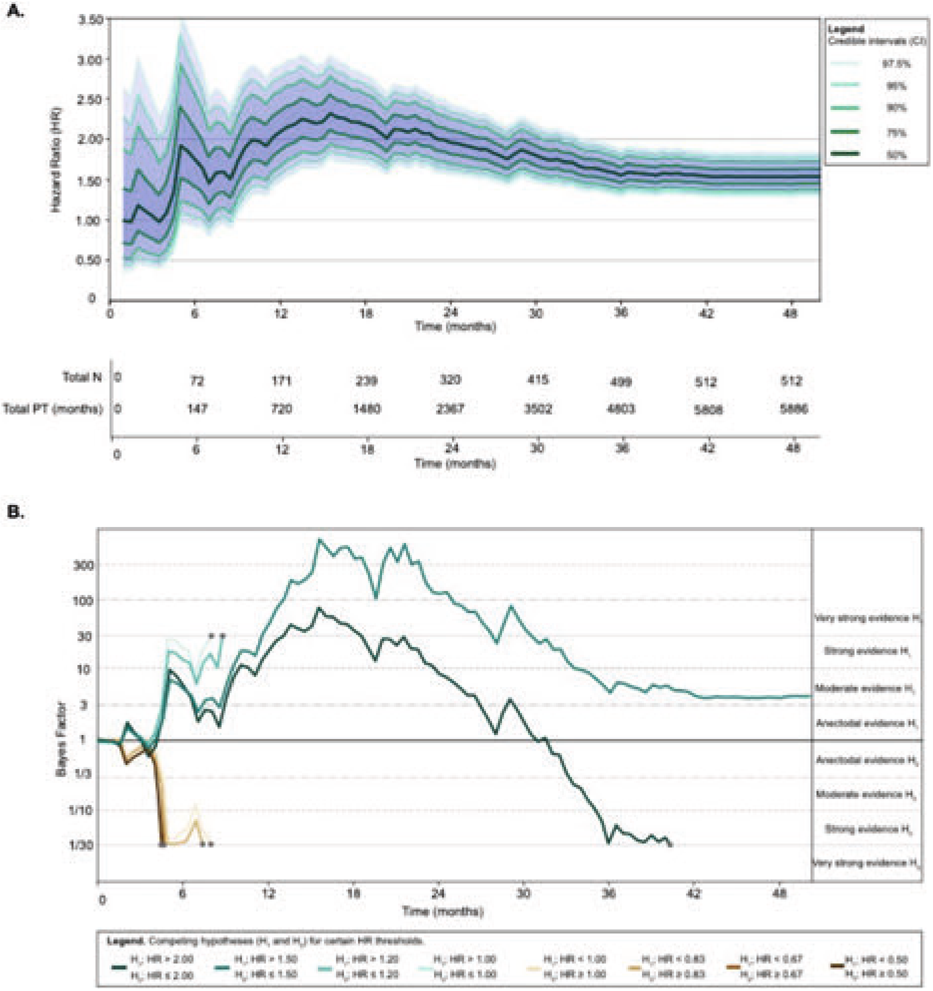
Figure 2b shows the of the Bayes Factor over time for different competing hypotheses, evaluating whether the observed HR falls above certain thresholds (ranging from HR >1.00 to >2.00) or below certain thresholds (ranging from HR <1.00 to <0.50). In the first 5 months data accumulation, there was strong to very strong evidence that the observed HR is above 1, indicating worse survival outcomes in clinical practice. At 15 months of data accumulation, there was very strong evidence that the observed HR is above 2. After 15 months of data accumulation, the strength of evidence for observing HRs above 2 decreased and when all data were included, there was strong evidence that the observed HR was above 1.0 and below 2.0, indicating that the survival of patients treated in clinical practice was worse than the survival for those treated in RCTs.

### Prior sensitivity analyses

The prior sensitivity analysis, detailed in Supplementary File S8 Figures S12 and S13 demonstrated that variations in the range of priors did not impact the results, confirming the robustness of the chosen prior. Similarly, the application of a different baseline hazard parameterization (using a Gompertz instead of a Weibull baseline hazard) did not affect the results, confirming robustness against baseline hazard parameterization (Supplementary File S9).

*Dataset 2: Survival outcomes associated with pembrolizumab monotherapy*

The results of the Bayesian survival model for dataset 2 are reported in Supplementary File S12. All model diagnostics were good, with posterior checks demonstrating a good fit between the model and observed data. After about 20 months of data accumulation, the model estimated precise HRs (Figure S20). When all data were included, the HR was 1.49 (95% credible interval 1.11 – 1.98) (Supplementary File S11, Table S6), indicating worse survival outcomes for patients treated in clinical practice than those treated in the RCT.



| | 0 | 6 | 12 | 18 | 24 | 30 | 36 | 42 | 48 |
|---|---|---|---|---|---|---|---|---|---|
| Total N | 0 | 72 | 171 | 239 | 320 | 415 | 499 | 512 | 512 |
| Total PT (months) | 0 | 147 | 720 | 1480 | 2367 | 3502 | 4803 | 5808 | 5886 |

**Figure 2.** Hazard Ratio (HR) represents the hazard of a survival event occurring in the clinical practice cohort divided by that in the RCT cohort for patients treated with pembrolizumab plus chemotherapy **A)** HR distribution over time with 50%, 75%, 95% and 97.5% credible intervals. **B)** The Bayes factor for competing hypotheses testing different HR thresholds for each model update. The grey circle indicates the point at which there is very strong evidence for either $H_1$ or $H_0$ for every further updated model.

## Discussion

We aimed to develop a Bayesian survival model that compares survival data from clinical practice that accumulates over time with static survival data from RCTs to investigate whether survival outcomes are similar, thereby providing rapid and easily interpretable results that can inform clinical and policy-related decision-making. Using two datasets comprising survival data from clinical practice and an RCT, we demonstrated that our Bayesian survival model enabled efficient comparisons of survival outcomes. For dataset 2, the model estimated precise HRs using part of the available data from clinical practice. The HR at the end of data accumulation was 1.49 for pembrolizumab monotherapy, indicating worse survival outcomes for patients treated in clinical practice than for those treated in the RCTs. In addition, these obtained estimates were equivalent to the ones obtained by traditional Cox PH regression modelling.[18] Furthermore, we demonstrated that sequential hypothesis testing with Bayes factors can be used for efficient and easily interpretable for results that can inform clinical and policy-related decision-making, with very strong evidence for the HR being > 1.0 and strong evidence for HR being > 1.2 before the end of data accumulation. For dataset 1, the posterior model checks showed a misfit between the model and the data after approximately 15 months, indicating that the data generation process was not well captured by the model. Thus, the model fit should be improved before reliable estimates of this dataset can be obtained. In dataset 1, we observed that the HR initially exceeded 2 in the first 15 months before decreasing to 1.54 by the end of data accumulation, suggesting worse survival outcomes for patients who started pembrolizumab plus chemotherapy shortly after it became available in clinical practice. This trend may reflect 'channeling' of novel treatments in clinical practice, where initially treated patients experience poorer outcomes due to use in more advanced disease or less optimized conditions. Over time, as protocols improve and clinicians gain experience, outcomes tend to improve.[21] To explore this phenomenon, another variable should be added to the model that explicitly captures the effect of treatment timing by linking the date of treatment initiation to its first availability. Importantly, this trend did not appear in the pembrolizumab monotherapy dataset, probably because patients receiving monotherapy in clinical practice represented a more homogenous population. As the first available first-line immunotherapy, pembrolizumab monotherapy was probably administered with more caution, leading to stricter patient selection than for pembrolizumab plus chemotherapy. Such variation in the data between the two datasets highlights the importance of rigorously monitoring posterior checks in Bayesian modelling, since these checks help to identify the model's limitations in capturing certain complexities in the data over time. Furthermore, further model developments are needed to improve the model fit in dataset 1, such as incorporating a variable that reflects the timing of treatment initiation relative to its availability.

Provided a good model fit – as with all models for statistical analysis – the Bayesian survival model developed in this study offers two key advantages: i) it facilitates earlier insights into potential differences in survival outcomes between accumulating clinical practice data and static RCT data, and ii) it facilitates easy interpretation of results to support clinical and policy-related decision-making. First, the model provided precise HR estimates for differences in survival outcomes between clinical practice and the RCT before including all data, indicating its potential to facilitate insights as early as statistically justifiable. Precise HR estimates were achieved approximately 10 months before the end of data accumulation (total 30 months) in dataset 2. As far as we are aware, no studies have applied a Bayesian approach for this specific context. In one study, a Bayesian survival model was developed for a post hoc analysis of two clinical trials with overall survival as the endpoint. The authors demonstrated that sequential Bayesian analyses provided earlier insights into treatment effects than frequentist Cox PH regression, with precise estimates available six months before the trial's data lock point.[22]

The second key advantage of the developed Bayesian approach is that it provides easily interpretable results to support clinical and policy-related decision-making. The model could generate early signals for treatment strategies that may be more or less effective in clinical practice than pre-specified HR thresholds specific to the type of medicine and clinical setting. For example, when the model shows a good posterior fit and there is very strong evidence that survival in clinical practice is substantially worse than in the trial (e.g. HR > 2), this could trigger further investigation into underlying causes and inform adjustments to treatment strategies. Currently, our Bayesian model incorporates only one variable (clinical practice versus RCT), providing HR estimates for the overall patient population rather than for specific subgroups. Ideally, additional variables, such as patient, disease, treatment, and setting variables, could be added to the model to provide direct insights for specific subgroups of patients. Adding other variables could also provide insights into how differences between clinical practice and RCT populations impact the observed HR estimates, as patients in clinical practice are often older, less fit, and more comorbid, with less standardized treatments and follow-up procedures. Although the model's explicit structure allows for this flexibility, the unavailability of individual patient data from RCTs limits the incorporation of such detailed characteristics for the current datasets. To still gain insight into the potential impact of patient and disease characteristics within the clinical practice population, we examined their distribution over time (Supplementary File S10 Tables S3 and S4). The distributions remained relatively stable, suggesting that changes in these characteristics are unlikely to explain the observed variation in HR over time.

In this study, the model has been applied to two datasets with different oncological treatments, investigating whether survival outcomes are similar for patients treated in clinical practice and those in RCTs. Beyond this application, the model can evaluate other time-to-event outcomes, such as time-to-treatment failure or time-to-next treatment, across various treatments and populations. Additionally, the model can compare time-to-event outcomes between two treatment strategies within clinical practice, offering valuable insights when head-to-head RCT data are unavailable. However, this requires further developments to the model to allow the incorporation of patients, disease and treatment variables.

Despite its advantages, our Bayesian survival model has several limitations. First, the use of a simple model with one covariate limited our ability to explore the impact of other covariates on the observed survival outcomes. However, as the study's primary focus was to design a proof-of-concept Bayesian framework to evaluate its feasibility and potential value in comparative effectiveness studies, a simple model was chosen to keep modeling concise and easily reproducible. However, it did not always provide an optimal fit, suggesting that other covariates may need to be screened for relevance and incorporated into the model to better fit the data. Additionally, the high level of censoring toward the end of the dataset suggests that parametric models may be less suitable, underscoring the need to investigate more flexible approaches, such as semi-parametric and non-parametric models.

## Conclusion

We developed a Bayesian survival model that compares survival data from clinical practice that accumulates over time with static survival data from RCTs to investigate whether survival outcomes are similar or different. Provided the posterior check shows an acceptable model fit, we demonstrated that sequential hypothesis testing with Bayes factors could be used to allow an accelerated route for informing clinical and policy-related decision-making. This approach provides a statistically sound method for stopping and inferring parameter estimates at the earliest possible time point (i.e., in contrast to the more traditional Cox modeling approach, where this is not possible because of practical and statistical concerns). The proposed visualizations allow for direct model interpretation by analyzing model fit and posterior confidence intervals around inferred parameter values. However, further developments to the model are needed to improve its ability to capture complexities in the data over time, such as incorporating a variable that captures the timing of treatment initiation relative to its availability. Additionally, incorporating patient, disease, and treatment characteristics into the model will facilitate investigating their impact on the observed survival outcomes.

## Funding

This study was supported by Roche Nederland BV and the Sint Antonius Research Fund.

## References

1. U.S. Food and Drug Administration. Clinical trial endpoints for the approval of cancer drugs and biologics: Guidance for industry. US Food and Drug Administration (2018).
2. European Medicine Agency. Committee for Human Medicinal Products ICH guideline E8 (R1) on general considerations for clinical studies (2021).
3. Feinstein AR, Horwitz MD. Double standards, scientific methods, and epidemiologic research. N Engl J of Med. 307(26):1611–1617 (1982).
4. Cramer-Van Der Welle CM, Peters BJM, Schramel FMNH, Klungel OH, Groen HJM, Van De Garde EMW. Systematic evaluation of the efficacy-effectiveness gap of systemic treatments in metastatic nonsmall cell lung cancer. Eur Respir J. 52(6):1801100 (2018).
5. Nordon C, Karcher H, Groenwold RHH, Ankarfeldt MZ, Pichler F, Chevrou-Severac H, et al. The "efficacy-effectiveness gap": Historical background and current conceptualization. Value in Health;19(1):75–81(2016).
6. Gebski V, Garès V, Gibbs E, Byth K. Data maturity and follow-up in time-to-event analyses. Int J Epidemiol. 47(3):850–9 (2018)
7. Brard C, Hampson L V, Gaspar N, Deley M cécile Le, Teuff G Le. Incorporating individual historical controls and aggregate treatment effect estimates into a Bayesian survival trial: a simulation study. *BMC Med Res Methodol*. 85: 1–17 (2019).
8. Bartoš F, Aust F, Haaf JM. Informed Bayesian survival analysis. *BMC Med Res Methodol*. 22(1):1–22 (2022).
9. van Rosmalen J, Dejardin D, van Norden Y, Löwenberg B, Lesaffre E. Including historical data in the analysis of clinical trials: Is it worth the effort? *Stat Methods Med Res*. 27(10):3167–82 (2018).
10. Ibrahim JG, Cheng MH, Sinha D. Why Bayes? Bayesian survival analysis. pp. 26-27. Springer Science & Business Media, 2013.
11. Kruschke JK. Bayesian analysis reporting guidelines. *Nat Hum Behav.* 5:1282-1291 (2021).
12. Zhao X, Yu C, Tong H. A Bayesian approach to Weibull survival model for clinical randomized censoring trial based on MCMC simulation. 2nd International Conference on Bioinformatics and Biomedical Engineering (iCBBE 2008). 4:1181-1184 (2008).
13. Omurlu IK, Ture M, Ozdamar K. Bayesian analysis of parametric survival models: A computer simulation study based on informative priors. *J Stat Manag Syst.* 18(5):405-423 (2015).
14. Paolucci I, Lin YM, Albuquerque Marques Silva J, Brock KK, Odisio BC. Bayesian parametric models for survival prediction in medical applications. *BMC Med Res Methodol.* 23(1):250 (2023).
15. Schönbrodt FD, Wagenmakers EJ, Zehetleitner M, Perugini M. Sequential hypothesis testing with Bayes factors: Efficiently testing mean differences. *Psychol Methods*. 22(2):322-339 (2017).
16. Gandhi L, Rodríguez-Abreu D, Gadgeel S, Esteban E, Felip E, De Angelis F, et al. Pembrolizumab plus chemotherapy in metastatic non–small-cell lung cancer. *N Engl J Med* 378(22):2078-2092 (2018).
17. Guyot P, Ades AE, Ouwens MJNM, Welton NJ. Enhanced secondary analysis of survival data: Reconstructing the data from published Kaplan-Meier survival curves. *BMC Med Res Methodol* 12:9 (2012).
18. Verschueren MV, Peters BJ, Bloem LT, Kruik VR, Uitvlugt EB, Bijsmans AR, et al. Pembrolizumab plus chemotherapy per PD-L1 stratum in patients with metastatic non–small cell lung cancer: Real-world effectiveness versus trial efficacy. *Clin Lung Cancer* 25(2):119-127 (2024).
19. Reck M, Rodríguez-Abreu D, Robinson AG, Hui R, Csoszi T, Fülöp A, et al. Updated analysis of KEYNOTE-024: Pembrolizumab versus platinum-based chemotherapy for advanced non–small-cell lung cancer with PD-L1 tumor proportion score of 50% or greater. *J Clin Oncol* 37(7):537-546 (2019).

20. Cramer-van der Welle CM, Verschueren MV, Tonn M, Peters BJM, Schramel FMNH, Klungel OH, et al. Real-world outcomes versus clinical trial results of immunotherapy in stage IV non-small cell lung cancer (NSCLC) in the Netherlands. *Sci Rep.* 11(1):23121 (2021).
21. Petri H, Urquhart J. Channeling bias in the interpretation of drug effects. *Stat Med* 10(4):577-581 (1991).
22. Biard L, Bergeron A, Lévy V, Chevret S. Bayesian survival analysis for early detection of treatment effects in phase 3 clinical trials. *Contemp Clin Trials Commun* 21:100709 (2021).

## Supplementary

Availability of data code**:**

- Project name: BayesianSurvival
- Project home page: https://github.com/danielverschueren/BayesianSurvival
- Programming language: Python

### Table of contents

## S1. Cox proportional hazard regression modeling

The Cox Proportional Hazards (PH) model is a widely used method to analyze time-to-event data, such as survival data, which measures the time until an event such as death. The Cox regression model is based on the hazard function and can be mathematically written as follows:

$$h(\boldsymbol{x}, t) = h_0(t)e^{\boldsymbol{x}\cdot\boldsymbol{\beta}} \quad (1)$$

The hazard function $h(\boldsymbol{x}, t)$ is the probability density that an event will occurring at time $t$, conditional on survival up until time $t$, given a set of $\kappa$ covariates $\boldsymbol{x} \in \{0,1\}^{\kappa}$. The baseline hazard, $h_0(t)$, represents the common hazard shared by all samples modified only by the covariates $\boldsymbol{x}$ and regression coefficients $\boldsymbol{\beta}$. By exponentiating the regression coefficients $\boldsymbol{\beta}$ the hazard ratio $HR$ can be obtained,

$$HR_{x(i)} = e^{\beta(i)}$$

for covariate $x(i)$ and with regression coefficient $\beta(i)$. It quantifies how each covariate influences the risk relative to the baseline. In our analysis, we incorporated only a single covariate $x$ representing the cohort type (observational vs RCT).

### Semi-parametric

The Cox PH is considered a semi-parametric model because it combines parametric and non-parametric components. The parametric component, represented by the regression coefficients $\boldsymbol{\beta} = (\beta(1), \beta(2), \ldots, \beta(\kappa))$ estimates the effect of covariates on the risk of an event. Meanwhile, in the Cox PH analysis the baseline hazard $h_0(t)$ does not need to be explicitly specified and is therefore non-parametric, providing generality and avoiding the necessity to assume a specific distribution for the timing of an event in the baseline model. This combination allows the model to adapt to complex data without making strong assumptions about the underlying risk pattern.

### Censoring

In most studies, not all patients experience the event within the study period, resulting in censored data where the exact event time is unknown. Right censoring is the most common form of censoring, occurring when the event not yet happened by the end of the follow-up period. Besides, censoring is often assumed to be non-informative, meaning that the reason for an individual being censored is unrelated to their risk of the event.

### Proportional Hazard

The Cox PH model, as exemplified by the constant-in-time coefficients $\boldsymbol{\beta}$ in Eqn. 1, relies on the proportional hazards assumption, meaning the hazard ratio between groups remains constant over time. In other words, the relative risk between two groups with different covariates is assumed to be the same at any point during the study period.

## S2. Bayesian survival estimation / model specifications

Bayesian analysis requires a parameterized likelihood function that describes the distribution of the observed quantity of interest, in our case the time-to-event $\boldsymbol{t}$. We start by defining a distribution of time-to-event $p_{event}(\boldsymbol{t}|\boldsymbol{m})$, for a given set of observed events. Our model parameters $\boldsymbol{m}$ specify the time-to-event distribution for which good (generative) models exist and we will keep them general for now. In survival analysis, it is convenient to express the time-to-event distribution in terms of the survival function $S(\boldsymbol{t}|\boldsymbol{m})$ and the *hazard rate* $h_0(\boldsymbol{t}|\boldsymbol{m})$, where

$$S(t|\boldsymbol{m}) = 1 - \int_0^t p_{event}(t'|\boldsymbol{m})dt' = \exp\left(-\int_0^t h(u|\boldsymbol{m})du\right),$$

and from which it follows that

$$p_{event}(t|\boldsymbol{m}) = h_0(t|\boldsymbol{m})S(t|\boldsymbol{m}).$$

Stating the time-to-event distribution in this way helps simplifying our analysis when subjects for which an event has *not yet* happened need to be included in the likelihood function.

The complete likelihood for observing a set of times $\boldsymbol{D} = \{t_0, t_1, ..\}$ with $\delta_i = 1$ for observing an event at time $\boldsymbol{t_i}$ or $\delta_i = 0$ for the absence of an event (survival) up until $\boldsymbol{t_i}$ then becomes:

$$p_{ll}(\boldsymbol{D}|\boldsymbol{m}) = \prod_i p_{event}(t_i|\boldsymbol{m})^{\delta_i} S(t_i|\boldsymbol{m})^{1-\delta_i}, (2)$$

which is the (baseline) likelihood function for survival we use for our modelling.

In our analysis, we are interested in the difference of survival probability between two sets of subjects, reference versus real world. Continuing, we will make a further assumption of *constant proportional hazards* (reiterating from Section S1):

$$h_\beta(t|\boldsymbol{m}, \beta) = h_0(t|\boldsymbol{m})e^\beta,$$

where the hazard rate of one set is constant proportional to the other by a factor

$$HR = e^{\beta},$$

with *HR* the *hazard ratio* and $\beta$ it's logarithm. $\beta > 0$ means an accelerated time-to-event distribution, $\beta < 0$ means a slowing of time-to-event distribution. In this instance, $h_0(t|\boldsymbol{m})$ is termed the baseline hazard and it is the hazard function for the reference set ($x_i = 0$) and $h_\beta(t|\boldsymbol{m}, \beta)$ is the hazard function for the real-world set ($x_i = 1$). Note that with the proportional hazards assumption, the time-to-event distribution for each set is still a proper distribution, as long as $h_0(t|\boldsymbol{m}) > 0$ for $t$. Generalizing Eqn. 2, the combined likelihood function for proportional hazards modelling then becomes:

$$p_{ll}(\boldsymbol{D}|\boldsymbol{m}, \beta) = \prod_i [h_0(t_i|\boldsymbol{m}) S(t_i|\boldsymbol{m})]^{\delta_i(1-x_i)} S(t_i|\boldsymbol{m})^{(1-\delta_i)(1-x_i)} \cdot$$

$$\left[h_\beta(t_i|\boldsymbol{m}) S(t_i|\boldsymbol{m})^{e^\beta}\right]^{\delta_i x_i} \left[S(t_i|\boldsymbol{m})^{e^\beta}\right]^{(1-\delta_i)x_i}. (3)$$

Fundamentally, we are interested in the parameter $\beta$, which quantifies the difference in survival between the two sets of subjects. The other parameters that define our baseline hazard $\boldsymbol{m}$ are not of direct interest to us and are our nuisance parameters.

In the Cox's proportional hazard's framework, the maximum-likelihood estimator can be evaluated such that the resulting value for $\beta$ becomes independent of the parameters $\boldsymbol{m}$ and hence need to be specified. However, in our Bayesian modelling, we do not have this option and need to explicitly define these parameters. Here, we take the time-to-event distribution to be of from the Weibull parametric family:

$$p_{event}(t|\boldsymbol{m}) = p_{event}(t|b, k) = bk(bt)^{k-1} e^{-(bt)^k},$$

with corresponding hazard function:

$$h_0(t|\boldsymbol{m}) = h(t|b, k) = bk(bt)^{k-1}.$$

The Weibull distribution is more flexible compared to the canonical exponential distribution $\lambda e^{-\lambda t}$. $b$ controls the scale of the distribution (similar to $\nu$ in the exponential distribution) and $k$ controls the increase ($k > 1$) or decrease ($k < 1$) of the hazard as subjects survive to later times. For $k = 1$ we recover the exponential distribution. The modeling as outlined here is general and any valid time-to-event distribution can be chosen, including more strongly parameterized versions where $h_0(t|\boldsymbol{m})$ consists for example of a set of restricted cubic-splines.

In all our modeling, the logarithm of Eqn. 3 is used to model the data, where the likelihood now depends on three explicit parameters: $p_u(\boldsymbol{D}|b, k, \beta)$ and we are interested in the posterior distribution of $\beta$: $p(\beta|\boldsymbol{D})$.

The priors for the various parameters $k, b$ and $\beta$ are defined as follows in our modelling:

- $k$ is taken from a uniform distribution with support $[0.5, 1.5]$.
- $b$ is taken from a LogNormal distribution, with parameters $\mu = -3, \sigma = 1$ to reflect the support on $[0, \infty]$.
- $\beta$ is taken as a Normal distribution with $\mu = 0, \sigma = 0.5$. Note that this implies an HR is log-normally distributed.

We have chosen these priors to include a reasonably wide range of models (see prior predictive checks) that data could have been generated by. The prior distributions for $\beta$, $b$, and $k$ are plotted in Figure S1.

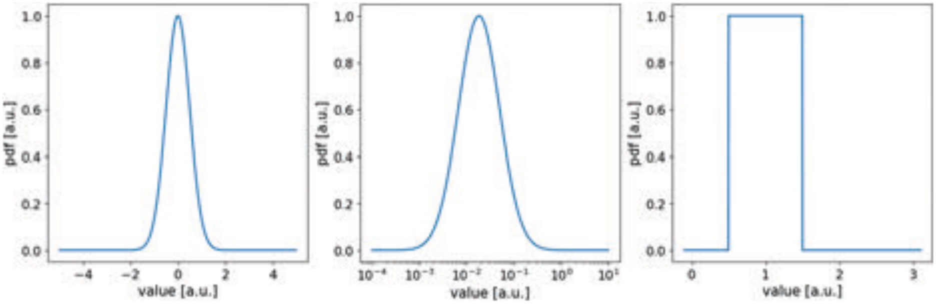


**Figure S1: Priors used for parameters in modelling, see text for details.** A: $\beta$, B: $b$, and C: $k$

We have performed prior-sensitivity tests to ensure our conclusions are not overly sensitive to our choice the prior distributions. The estimation of the parameters $b, k$ are insensitive to our choice of prior, as they are already from the first time point heavily influenced by all the data from the (large) reference set. In case this set is also built up as time develops, the choice of prior for these parameters can be more influential to the results.

## S3. Bayesian survival model applied to synthetic data

To validate the Bayesian survival model, we applied the model to synthetically generated survival data. We used the following generative model to create the data:

$$p(t_j|\boldsymbol{m}) \sim Exponential(t_i|\, \nu = \nu_i)$$

for every observable $t_j : t_{ref}, t_{end}, t_{cens}$, with parameters $\nu_{ref} = 0.05$, $\nu_{cens} = 0.1$ and $\nu_{end} = \nu_{ref} \cdot \mathrm{e}^{\beta} = 0.05e^{0.5}$ ($HR = 1.64$, and consistent with the Cox Proportional Hazard assumption for an exponential baseline hazard). The independent exponential chosen to generate censored time values approximates a censoring process that consists of the amalgamation of processes that determine whether a subject will drop out from the synthetic study. For each reference sample, an event time $t_{end}$, and a censoring time $t_{cens}$ is drawn, akin to the experimental setup under study in this work ($t_{start} = 0$ for all samples) and only a reference time $t_{ref}$ is drawn for the test arm. Between $t_{end}$ and $t_{cens}$, the smallest value is chosen, per simulated sample. The sample is then assigned to the proper class, event or censored, mirroring a natural process where the censoring process competes with the survival process under study. The relative probability of a data sample being drawn from either the event distribution $t_{end}$ or the censoring distribution $t_{cens}$ is thus implicitly determined and not strictly controlled. In this configuration, about 50% of the subject are censored.

After drawing 100 samples from this model, we applied our Bayesian survival model using an exponential baseline hazard rate to infer the proper baseline hazard $\nu_{ref}$ and proportional hazard $\beta$, using a prior LogNormal distribution over $\nu_{ref}$, with parameters $\mu = -3, \sigma = 1$ and the prior for $\beta$ a Normal distribution with $\mu = 0, \sigma = 0.5$. As is clear from the results shown in Fig. S2 below, the model infers the correct values for $\nu_{ref}$ and $\beta$. We have added this simple dataset to the source code repository of our Bayesian model.

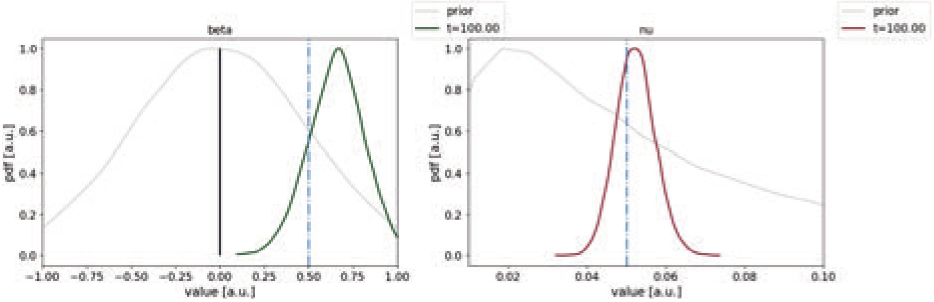


**Figure S2: Posteriors for synthetic dataset.** A: $\beta$, B: $\nu_{\mathrm{ref}}$. The parameters are correctly inferred by the model $\beta = 0.74 \pm 0.35$ c.f. $0.5$, $\nu = 0.053 \pm 0.16$ c.f. $0.05$.

## S4. Posterior MCMC diagnostics

The pymc MCMC sampler was set up using, 4 chains of 5000 samples, a warm-up length of 6000 samples, and an acceptance rate of 0.9. A single posterior MCMC optimization of the posteriors would take approximately 0.5 min for both dataset 1 and dataset 2 (single chain, single 2.3GHz Intel CPU (Intel MacBook Pro 2020)).

For all posterior fits, the chains demonstrated excellent convergence with inter-to-intra chain variability abs(rhat - 1) < 0.001 and effective number of samples (ess) > 5500 for all parameters $\beta$, $b$, and $k$. Table S1 and S2 show the effective number of samples for each parameter, per MCMC run. The figure S3 and S4 shows a typical set of traces for all chains of the MCMC solution on the posteriors for all parameters for dataset 1 (Figure S3) and dataset 2 (Figure S4). The absence of drift in the chains as the number of drawn samples progresses and the agreement between the chains indicate excellent convergence.

At every time point of evaluation, when new data have come available, we have rerun the complete MCMC posterior calculation starting from the prior distribution and including all data available up until that point, rather than using the posterior from the previous timepoint and updating it with only the new data. However, in a Bayesian setting, these two approaches are similar up to MCMC sampling error.

**Table S1.** effective samples (ess) per MCMC run and inter-to-intra chain variability (r-hat) for dataset 1.

| T | ESS_B | ESS_K | ESS_BETA | RHAT_B | RHAT_K | RHAT_BETA |
|---|---|---|---|---|---|---|
| **1** | 7565.796 | 7480.536 | 10196.349 | 1.000 | 1.000 | 1.000 |
| **1.5** | 7394.666 | 7597.070 | 10206.295 | 1.000 | 1.000 | 1.000 |
| **2** | 7476.803 | 7376.973 | 10772.897 | 1.000 | 1.001 | 1.000 |
| **2.5** | 7621.060 | 7664.517 | 9529.799 | 1.000 | 1.001 | 1.000 |
| **3** | 8102.871 | 7979.867 | 10224.610 | 1.000 | 1.000 | 1.000 |
| **3.5** | 8204.111 | 8162.187 | 10023.622 | 1.000 | 1.000 | 1.000 |
| **4** | 8752.948 | 8851.309 | 10136.079 | 1.001 | 1.001 | 1.000 |
| **4.5** | 7023.677 | 6880.936 | 9385.681 | 1.000 | 1.000 | 1.001 |
| **5** | 6873.145 | 6880.853 | 9826.018 | 1.000 | 1.000 | 1.000 |
| **5.5** | 7326.588 | 7325.889 | 8790.643 | 1.001 | 1.000 | 1.000 |
| **6** | 6872.818 | 6893.629 | 9719.481 | 1.000 | 1.000 | 1.000 |
| **6.5** | 6982.499 | 7024.829 | 8808.045 | 1.000 | 1.000 | 1.000 |
| **7** | 7524.579 | 7554.512 | 8228.358 | 1.001 | 1.001 | 1.001 |
| **7.5** | 7097.302 | 7116.485 | 8801.326 | 1.000 | 1.000 | 1.000 |
| **8** | 6614.765 | 6800.607 | 8837.995 | 1.000 | 1.000 | 1.000 |
| **8.5** | 7522.511 | 7625.203 | 8804.207 | 1.001 | 1.001 | 1.000 |
| **9** | 7389.140 | 7554.582 | 8363.300 | 1.001 | 1.001 | 1.001 |
| **9.5** | 7339.292 | 7441.921 | 8002.261 | 1.000 | 1.000 | 1.000 |
| **10** | 6026.702 | 6113.212 | 8315.664 | 1.001 | 1.001 | 1.000 |
| **10.5** | 5996.768 | 6125.984 | 8420.336 | 1.001 | 1.001 | 1.001 |
| **11** | 6426.333 | 6553.721 | 7363.560 | 1.000 | 1.000 | 1.000 |
| **11.5** | 6334.405 | 6417.945 | 7127.455 | 1.001 | 1.001 | 1.000 |
| **12** | 6151.495 | 6216.021 | 7525.287 | 1.001 | 1.001 | 1.001 |
| **12.5** | 6479.212 | 6681.981 | 7388.955 | 1.000 | 1.000 | 1.000 |
| **13** | 6010.072 | 6367.218 | 7209.013 | 1.000 | 1.000 | 1.000 |
| **13.5** | 5959.008 | 6316.286 | 6976.178 | 1.001 | 1.001 | 1.001 |
| **14** | 6805.649 | 7072.552 | 7234.893 | 1.000 | 1.000 | 1.001 |
| **14.5** | 7221.343 | 7542.397 | 7835.050 | 1.001 | 1.001 | 1.001 |
| **15** | 6188.850 | 6478.072 | 7268.782 | 1.001 | 1.001 | 1.001 |
| **15.5** | 6412.511 | 6560.610 | 7792.380 | 1.002 | 1.002 | 1.001 |
| **16** | 6181.078 | 6517.821 | 7748.258 | 1.001 | 1.001 | 1.000 |
| **16.5** | 5956.285 | 6308.017 | 6990.152 | 1.001 | 1.001 | 1.000 |
| **17** | 6373.086 | 6722.937 | 7823.779 | 1.001 | 1.001 | 1.001 |
| **17.5** | 5761.054 | 5839.300 | 8327.538 | 1.000 | 1.001 | 1.000 |
| **18** | 6557.457 | 6756.855 | 7948.428 | 1.001 | 1.001 | 1.000 |
| **18.5** | 6056.646 | 6004.620 | 6986.261 | 1.001 | 1.001 | 1.000 |
| **19** | 6570.613 | 6742.065 | 7726.927 | 1.000 | 1.001 | 1.000 |

**Table S1.** Continued.

| T | ESS_B | ESS_K | ESS_BETA | RHAT_B | RHAT_K | RHAT_BETA |
|---|---|---|---|---|---|---|
| **19.5** | 6671.466 | 6956.651 | 8007.651 | 1.000 | 1.000 | 1.000 |
| **20** | 7120.439 | 7416.474 | 7534.766 | 1.000 | 1.000 | 1.001 |
| **20.5** | 6597.914 | 6876.292 | 7704.385 | 1.000 | 1.000 | 1.000 |
| **21** | 6140.314 | 6502.693 | 7518.814 | 1.001 | 1.001 | 1.000 |
| **21.5** | 6069.262 | 6359.112 | 7475.389 | 1.001 | 1.001 | 1.001 |
| **22** | 6438.490 | 6556.801 | 8446.919 | 1.001 | 1.001 | 1.001 |
| **22.5** | 6617.772 | 6567.010 | 8120.425 | 1.000 | 1.000 | 1.000 |
| **23** | 6720.428 | 7046.241 | 7882.419 | 1.001 | 1.001 | 1.000 |
| **23.5** | 6234.148 | 6566.225 | 7683.184 | 1.000 | 1.001 | 1.000 |
| **24** | 6766.989 | 7071.521 | 8140.774 | 1.001 | 1.000 | 1.000 |
| **24.5** | 6031.678 | 6207.259 | 8029.286 | 1.000 | 1.000 | 1.001 |
| **25** | 5890.908 | 6094.922 | 8322.240 | 1.000 | 1.000 | 1.000 |
| **25.5** | 7138.341 | 7196.569 | 8694.102 | 1.001 | 1.000 | 1.000 |
| **26** | 6311.400 | 6697.248 | 8153.973 | 1.001 | 1.000 | 1.000 |
| **26.5** | 7226.028 | 7593.112 | 8597.330 | 1.001 | 1.001 | 1.000 |
| **27** | 5742.691 | 6046.454 | 7443.544 | 1.001 | 1.001 | 1.000 |
| **27.5** | 6875.656 | 7377.381 | 8326.725 | 1.000 | 1.000 | 1.000 |
| **28** | 6358.361 | 6730.277 | 7757.103 | 1.001 | 1.000 | 1.000 |
| **28.5** | 6302.896 | 6983.513 | 7680.664 | 1.000 | 1.000 | 1.001 |
| **29** | 6440.024 | 6928.280 | 8089.078 | 1.000 | 1.000 | 1.001 |
| **29.5** | 6681.213 | 7044.400 | 8304.672 | 1.001 | 1.001 | 1.001 |
| **30** | 7638.401 | 8084.144 | 8461.265 | 1.000 | 1.000 | 1.000 |
| **30.5** | 6786.306 | 7211.516 | 8064.901 | 1.000 | 1.000 | 1.000 |
| **31** | 7089.433 | 7396.837 | 7548.310 | 1.001 | 1.000 | 1.000 |
| **31.5** | 7083.359 | 7324.931 | 8455.707 | 1.000 | 1.000 | 1.000 |
| **32** | 6771.812 | 7178.189 | 8198.438 | 1.001 | 1.001 | 1.000 |
| **32.5** | 6108.976 | 6584.236 | 7907.865 | 1.001 | 1.001 | 1.001 |
| **33** | 7399.232 | 7676.615 | 9218.894 | 1.000 | 1.000 | 1.000 |
| **33.5** | 6634.854 | 7031.417 | 8146.025 | 1.001 | 1.001 | 1.001 |
| **34** | 6335.861 | 6766.154 | 8033.736 | 1.001 | 1.001 | 1.000 |
| **34.5** | 7353.552 | 7872.624 | 8227.362 | 1.000 | 1.000 | 1.000 |
| **35** | 6955.490 | 7245.490 | 8426.351 | 1.000 | 1.000 | 1.000 |
| **35.5** | 6222.710 | 6669.901 | 8468.150 | 1.000 | 1.000 | 1.001 |
| **36** | 6890.845 | 7137.179 | 8557.841 | 1.000 | 1.000 | 1.001 |
| **36.5** | 6654.836 | 7057.359 | 8310.517 | 1.001 | 1.001 | 1.000 |
| **37** | 6543.993 | 6743.802 | 9025.348 | 1.001 | 1.001 | 1.000 |
| **37.5** | 6702.435 | 7142.315 | 8319.555 | 1.001 | 1.000 | 1.001 |
| **38** | 6645.793 | 7008.216 | 8113.259 | 1.001 | 1.000 | 1.000 |

**Table S1.** Continued.

| T | ESS_B | ESS_K | ESS_BETA | RHAT_B | RHAT_K | RHAT_BETA |
|---|---|---|---|---|---|---|
| **38.5** | 6796.773 | 7177.277 | 8220.345 | 1.001 | 1.000 | 1.001 |
| **39** | 6673.952 | 7269.500 | 7957.445 | 1.000 | 1.000 | 1.000 |
| **39.5** | 7189.971 | 7595.818 | 9179.917 | 1.001 | 1.000 | 1.000 |
| **40** | 6960.847 | 7300.348 | 8301.402 | 1.001 | 1.001 | 1.001 |
| **40.5** | 6283.693 | 6796.706 | 8337.499 | 1.001 | 1.000 | 1.000 |
| **41** | 6334.968 | 6578.484 | 8621.104 | 1.000 | 1.000 | 1.000 |
| **41.5** | 6386.011 | 6729.294 | 8847.848 | 1.000 | 1.000 | 1.000 |
| **42** | 7350.646 | 7977.102 | 7968.636 | 1.001 | 1.000 | 1.001 |
| **42.5** | 6915.257 | 7259.648 | 8665.718 | 1.001 | 1.000 | 1.000 |
| **43** | 6585.701 | 7515.628 | 8484.134 | 1.000 | 1.000 | 1.000 |
| **43.5** | 6325.064 | 6809.310 | 8605.331 | 1.000 | 1.000 | 1.001 |
| **44** | 7199.194 | 7614.194 | 8717.326 | 1.001 | 1.000 | 1.000 |
| **44.5** | 6825.473 | 7083.190 | 8919.228 | 1.000 | 1.000 | 1.001 |
| **45** | 7722.821 | 7985.734 | 8949.496 | 1.000 | 1.000 | 1.000 |
| **45.5** | 7503.021 | 8201.712 | 8025.572 | 1.000 | 1.000 | 1.000 |
| **46** | 6593.713 | 7277.000 | 8868.172 | 1.000 | 1.000 | 1.000 |
| **46.5** | 7055.226 | 7637.006 | 7931.607 | 1.000 | 1.000 | 1.001 |
| **47** | 7067.795 | 7549.208 | 8518.981 | 1.000 | 1.000 | 1.000 |
| **47.5** | 6777.020 | 7342.591 | 8503.309 | 1.000 | 1.000 | 1.000 |
| **48** | 7971.619 | 8338.347 | 8750.097 | 1.000 | 1.000 | 1.000 |
| **48.5** | 7341.767 | 7971.602 | 8785.617 | 1.000 | 1.000 | 1.001 |
| **49** | 7341.767 | 7971.602 | 8785.617 | 1.000 | 1.000 | 1.001 |
| **49.5** | 7341.767 | 7971.602 | 8785.617 | 1.000 | 1.000 | 1.001 |
| **50** | 7341.767 | 7971.602 | 8785.617 | 1.000 | 1.000 | 1.001 |

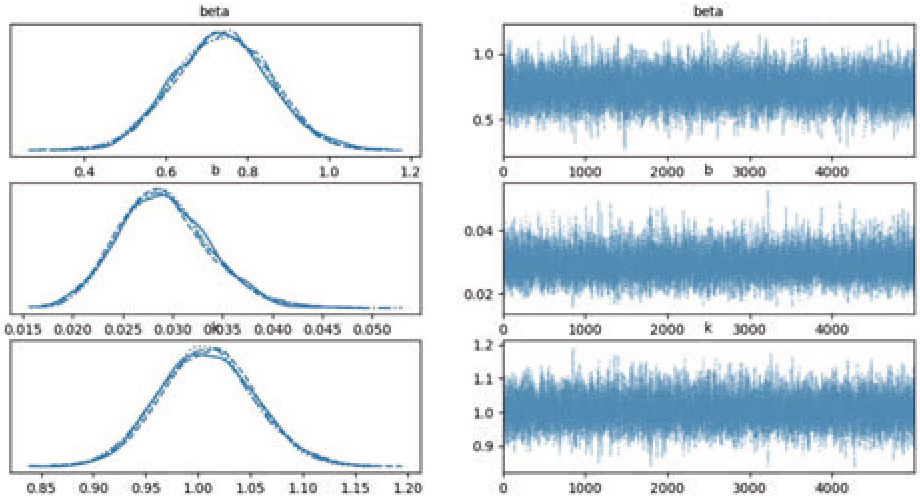


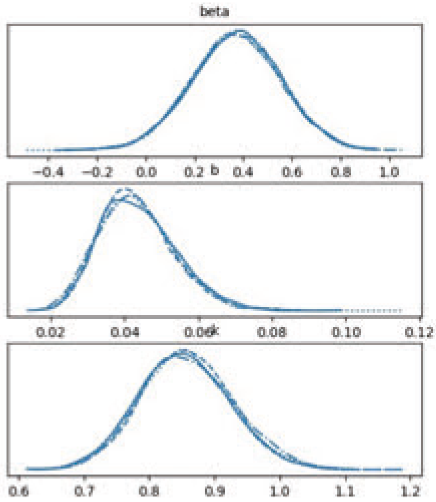
**Figure S3: Example of MCMC convergence of chains (6000 warm-up + 4 chains of 5000 samples) for dataset 1 at T=22 months.** The kernel-density plot for each chain for each parameter is plotted on the left-hand of the figure, the samples as they are drawn from the sampler in time are shown on the right. The absence of drift (left) and agreement between the kernel-density plots on the posteriors (right) for each sample and chain indicate good convergence.

**Table S2.** effective samples (ess) per MCMC run and inter-to-intra chain variability (r-hat) for dataset 2.

| T | ESS_B | ESS_K | ESS_BETA | RHAT_B | RHAT_K | RHAT_BETA |
|---|---|---|---|---|---|---|
| **1** | 8083.523 | 7972.556 | 10746.032 | 1.000 | 1.000 | 1.001 |
| **1.5** | 7823.346 | 7596.797 | 10533.612 | 1.000 | 1.000 | 1.000 |
| **2** | 7786.460 | 7688.818 | 10447.288 | 1.001 | 1.001 | 1.001 |
| **2.5** | 7870.122 | 7896.283 | 10260.419 | 1.000 | 1.000 | 1.001 |
| **3** | 6910.747 | 6912.016 | 10022.010 | 1.001 | 1.001 | 1.000 |
| **3.5** | 7197.176 | 7143.812 | 9512.787 | 1.001 | 1.001 | 1.000 |
| **4** | 6753.870 | 6537.132 | 11246.688 | 1.000 | 1.001 | 1.000 |
| **4.5** | 7303.096 | 7222.573 | 10150.853 | 1.001 | 1.001 | 1.000 |
| **5** | 6719.080 | 6859.570 | 10323.368 | 1.000 | 1.000 | 1.000 |
| **5.5** | 6947.894 | 6713.617 | 10532.862 | 1.000 | 1.000 | 1.000 |
| **6** | 7097.473 | 7058.722 | 11014.704 | 1.000 | 1.001 | 1.000 |
| **6.5** | 7966.885 | 7860.419 | 8937.987 | 1.000 | 1.000 | 1.000 |
| **7** | 7628.681 | 7525.762 | 9711.456 | 1.001 | 1.000 | 1.000 |
| **7.5** | 6486.383 | 6489.513 | 9531.953 | 1.001 | 1.001 | 1.000 |
| **8** | 8066.181 | 8197.518 | 9468.216 | 1.001 | 1.001 | 1.001 |
| **8.5** | 7370.456 | 7343.084 | 9263.877 | 1.001 | 1.000 | 1.000 |
| **9** | 6501.705 | 6623.510 | 9418.344 | 1.000 | 1.000 | 1.000 |
| **9.5** | 7965.230 | 7965.257 | 9891.900 | 1.001 | 1.001 | 1.001 |
| **10** | 6399.765 | 6446.942 | 9329.065 | 1.001 | 1.000 | 1.000 |
| **10.5** | 6981.500 | 7072.846 | 9735.262 | 1.000 | 1.000 | 1.000 |
| **11** | 7499.797 | 7689.796 | 8626.474 | 1.000 | 1.001 | 1.001 |
| **11.5** | 6617.425 | 6623.685 | 9076.428 | 1.001 | 1.001 | 1.001 |
| **12** | 6595.352 | 6736.690 | 8746.684 | 1.000 | 1.000 | 1.000 |
| **12.5** | 5478.191 | 5811.496 | 9333.600 | 1.000 | 1.001 | 1.000 |
| **13** | 6299.195 | 6449.544 | 9083.873 | 1.000 | 1.000 | 1.000 |
| **13.5** | 6830.154 | 6679.192 | 8781.348 | 1.001 | 1.001 | 1.001 |
| **14** | 6964.695 | 7169.103 | 8154.403 | 1.000 | 1.000 | 1.001 |
| **14.5** | 6843.042 | 6982.227 | 8249.098 | 1.000 | 1.000 | 1.001 |
| **15** | 5706.426 | 5794.868 | 8906.072 | 1.001 | 1.001 | 1.001 |
| **15.5** | 6603.253 | 6918.820 | 8608.695 | 1.001 | 1.001 | 1.000 |
| **16** | 6321.597 | 6679.037 | 8467.141 | 1.001 | 1.000 | 1.001 |
| **16.5** | 6462.948 | 6534.813 | 8861.702 | 1.001 | 1.000 | 1.001 |
| **17** | 6381.808 | 6587.491 | 8652.296 | 1.000 | 1.000 | 1.000 |
| **17.5** | 6886.390 | 7104.169 | 8539.432 | 1.000 | 1.000 | 1.000 |
| **18** | 6145.415 | 6739.088 | 8826.982 | 1.001 | 1.001 | 1.000 |
| **18.5** | 6581.353 | 6674.725 | 9230.157 | 1.001 | 1.000 | 1.000 |
| **19** | 6062.297 | 6260.975 | 8448.790 | 1.001 | 1.000 | 1.000 |

**Table S2:** Continued.

| T | ESS_B | ESS_K | ESS_BETA | RHAT_B | RHAT_K | RHAT_BETA |
|---|---|---|---|---|---|---|
| **19.5** | 6211.982 | 6497.891 | 7941.316 | 1.001 | 1.001 | 1.000 |
| **20** | 7328.317 | 7470.276 | 8835.332 | 1.000 | 1.000 | 1.000 |
| **20.5** | 7265.093 | 7683.219 | 8722.013 | 1.001 | 1.000 | 1.001 |
| **21** | 6311.527 | 6680.668 | 8181.408 | 1.000 | 1.000 | 1.001 |
| **21.5** | 6489.090 | 6855.840 | 8633.257 | 1.000 | 1.000 | 1.001 |
| **22** | 7221.132 | 7419.992 | 9230.492 | 1.000 | 1.000 | 1.001 |
| **22.5** | 6751.173 | 7054.237 | 8635.506 | 1.000 | 1.000 | 1.001 |
| **23** | 7815.349 | 7989.221 | 8804.778 | 1.000 | 1.001 | 1.000 |
| **23.5** | 6485.889 | 6940.608 | 8527.564 | 1.000 | 1.000 | 1.000 |
| **24** | 7710.328 | 7787.556 | 9518.326 | 1.001 | 1.000 | 1.001 |
| **24.5** | 7436.914 | 7500.156 | 9620.549 | 1.000 | 1.000 | 1.000 |
| **25** | 6146.401 | 6794.192 | 8344.721 | 1.001 | 1.001 | 1.000 |
| **25.5** | 7612.847 | 7707.056 | 9545.294 | 1.000 | 1.000 | 1.001 |
| **26** | 7283.577 | 7344.893 | 9128.746 | 1.000 | 1.000 | 1.000 |
| **26.5** | 6793.392 | 6921.465 | 9825.125 | 1.000 | 1.000 | 1.000 |
| **27** | 7654.356 | 7745.579 | 9734.094 | 1.000 | 1.001 | 1.000 |
| **27.5** | 7577.436 | 7772.487 | 10156.000 | 1.000 | 1.001 | 1.000 |
| **28** | 7249.352 | 7500.605 | 9737.662 | 1.001 | 1.001 | 1.001 |
| **28.5** | 7556.421 | 7682.136 | 9766.366 | 1.001 | 1.001 | 1.000 |
| **29** | 8114.242 | 8298.894 | 9459.684 | 1.001 | 1.000 | 1.001 |
| **29.5** | 8026.250 | 8401.685 | 10169.125 | 1.000 | 1.000 | 1.001 |
| **30** | 7270.351 | 7844.307 | 9257.223 | 1.000 | 1.000 | 1.000 |
| **30.5** | 7985.869 | 8326.498 | 9695.012 | 1.000 | 1.000 | 1.000 |
| **31** | 7985.869 | 8326.498 | 9695.012 | 1.000 | 1.000 | 1.000 |
| **31.5** | 7985.869 | 8326.498 | 9695.012 | 1.000 | 1.000 | 1.000 |
| **32** | 7985.869 | 8326.498 | 9695.012 | 1.000 | 1.000 | 1.000 |



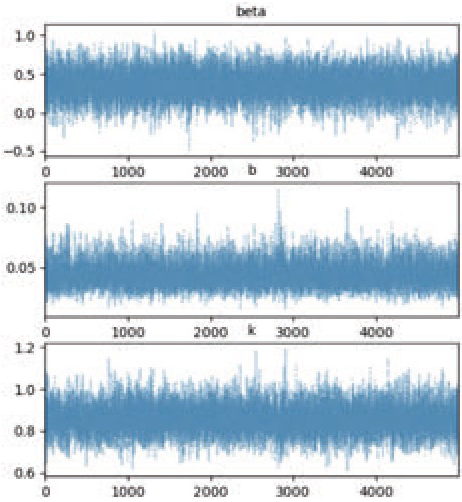
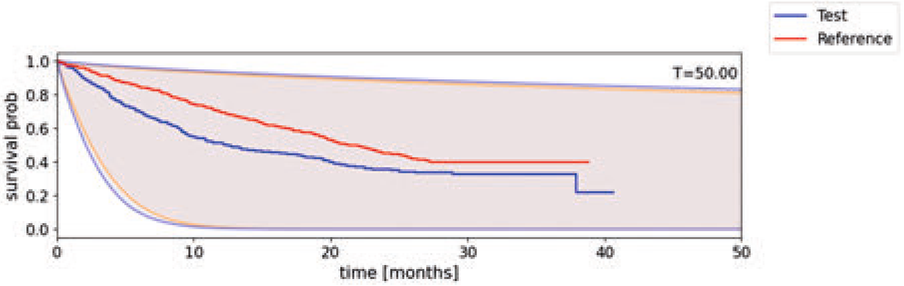


**Figure S4: Example of MCMC convergence of chains (6000 warm-up + 4 chains of 5000 samples) for dataset 2 at T=17 months.** The kernel-density plot for each chain for each parameter is plotted on the left-hand of the figure, the samples as they are drawn from the sampler in time are shown on the right. The absence of drift (left) and agreement between the kernel-density plots on the posteriors (right) for each sample and chain indicate good convergence.

## S5. Prior predictive checks



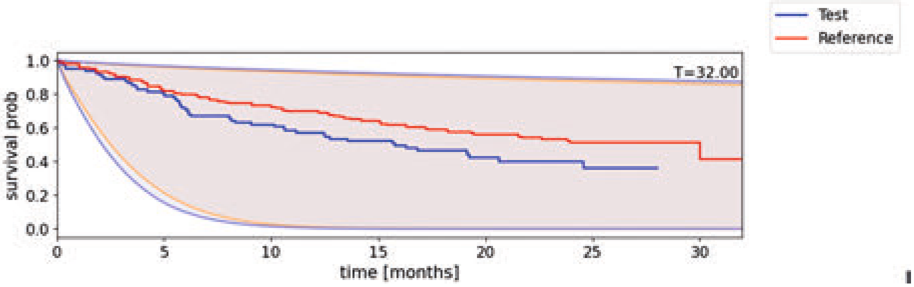


**Figure S5: Prior check for both reference (trial data, red) and test (real world data, blue) parameters, where each shaded area represents the [5%,95%] credible interval (orange: reference, purple: test) for the model drawn from the priors.** The resulting models span a large range of possible observed life-time distributions. The reference and test priors span both a wide range as a result of the wide prior on b and the exponential decay in lifetime distribution times. Top: dataset 1, bottom: dataset 2.

## S6. Posterior predictive checks

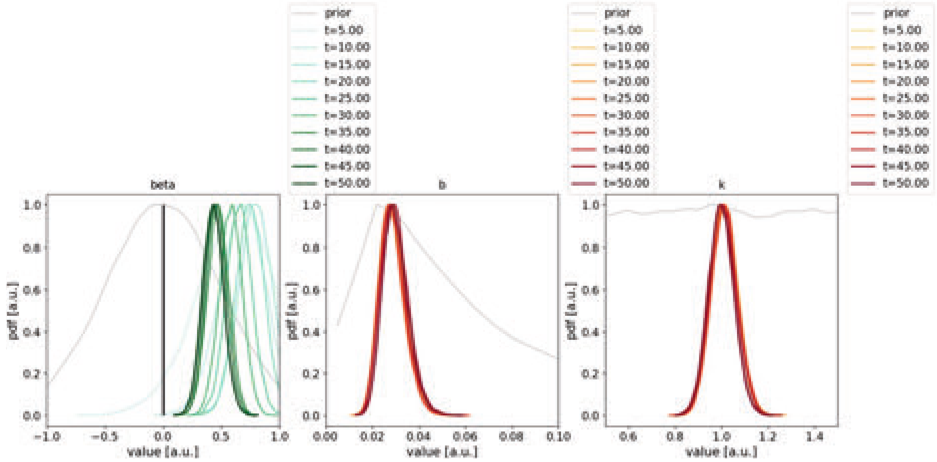


**Figure S6: Posterior checks for dataset 1.** Evolution of posterior for (universal) time $T = 5, \ldots, 50$ months in intervals of 5 months for parameters $\beta$, $b$, and $k$.

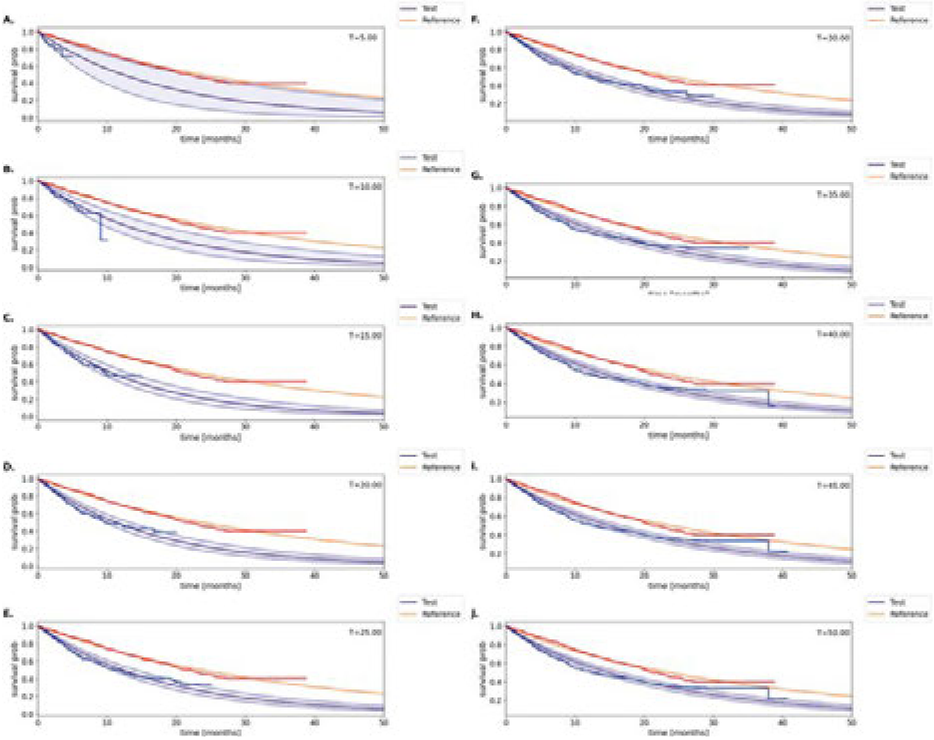


**Figure S7: Posterior check model fits to reference (trial, red) and test (real world, blue) data from dataset 1, where each shaded area represents the [5%,95%] credible interval on the test survival models drawn from the posterior (purple).** The solid orange and purple curves represent the median model drawn from the posterior for the reference and test sets. The credible around the test set narrow as time progresses. Up until $\mathrm{T} = 15$ test data indicate a good overlap with the posterior model, but thereafter there is a consistent misfit between data and model manifests. This could indicate a poor baseline hazard model or a shifting of the population of test data over time. The latter is most likely, as illustrated also by the strong movement of the test beta posterior over time and has been reported in clinical practice for real world cohorts focusing on oncological therapies. (4)

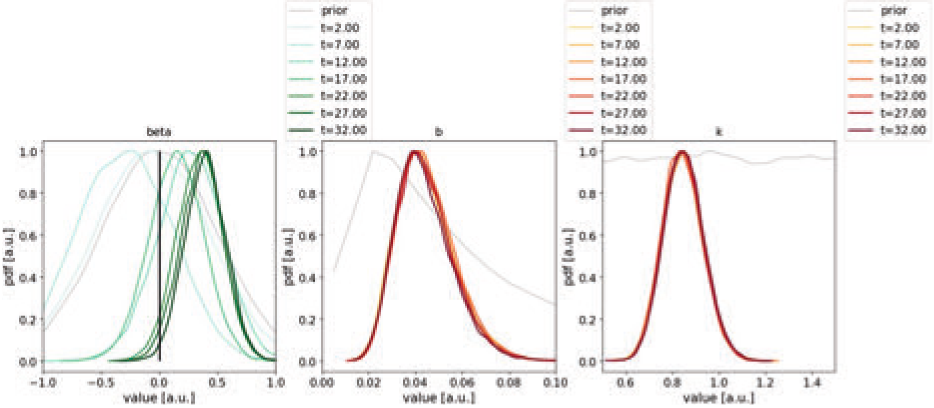


**Figure S8: Posterior checks for dataset 2.** Evolution of posterior for (universal) time $T = 2, \ldots, 32$ months in intervals of 5 months for parameters $\beta$, $b$, and $k$.

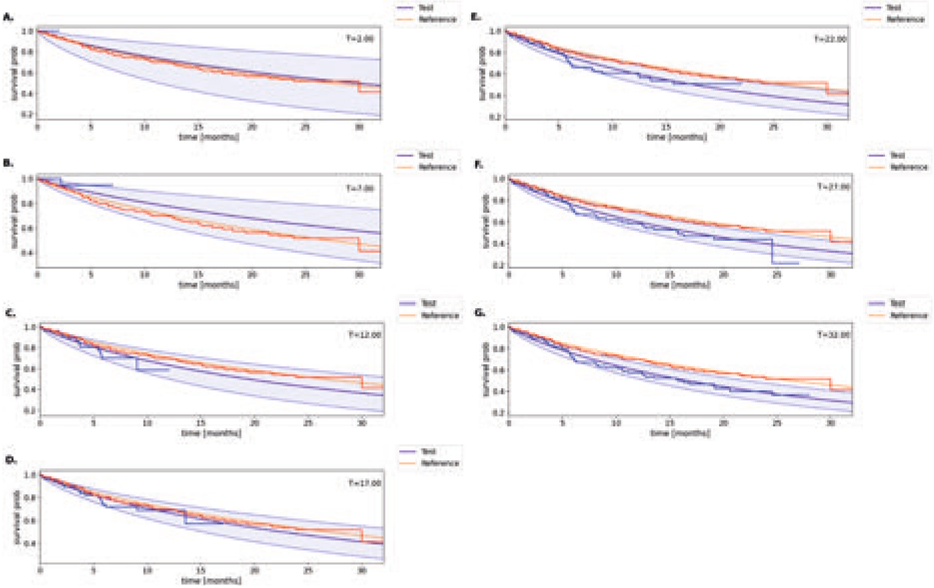


**Figure S9: Posterior check model fits to reference (trial, red) and test (real world, blue) data from dataset 2, where each shaded area represents the [5%,95%] credible interval on the test survival models drawn from the posterior (purple).** The solid orange and purple curves represent the median model drawn from the posterior for the reference and test sets. The credible around the test set narrow as time progresses and the test data lies comfortably in the model credible span, indicating good fits and a clear separation between reference and test model appears.

## S7. Sequential Bayes Factor hypothesis testing

In traditional null hypothesis testing, the probability of observing a certain value for the estimator of a decision score under a null model $H_0$ is evaluated. In case the probability $p$ of observing an estimated value is smaller than a certain threshold value $\alpha$, commonly $0.05$, $H_0$ is rejected, the data are decided to *not* have been generated by this model, and action is taken. In this framework, the probability of erroneously rejecting the $H_0$ (type I error) is directly controlled and an interpretation of significance can be ascribed to the decision.

However, this framework has a couple of drawbacks. First, corrections need to be applied to the threshold value $\alpha$ when multiple tests are performed on the same (or partially overlapping) data, as by definition $H_0$ is expected to be rejected erroneously once in every $1/p$ times the test is evaluated. Second, the alternative hypothesis $H_1$ is not explicitly defined and might be equally unlikely for the observed data, rendering its acceptance tenuous. (1)

When adopting a Bayesian framework, we can mitigate these two drawbacks by directly evaluating the relative *hypothesis evidence* $p(H_i|\boldsymbol{D})$ for both the null and alternative hypotheses and make decisions based on a Bayes factor.

$$BF_{01} = \frac{p(H_1|\boldsymbol{D})}{p(H_0|\boldsymbol{D})} = \frac{p(H_1)p(\boldsymbol{D}|H_1)}{p(H_0)p(\boldsymbol{D}|H_0)} \ .$$

Here, $\boldsymbol{D}$ is the observed data and $p(\boldsymbol{D}|H_i)$ is the marginal of the posterior (also called the evidence) over all possible model parameters specified in $H_i$. Usually, the priors $p(H_i)$ for both hypotheses are taken to be the same. The Bayes factor can be interpreted as a direct quantification of the favoring of hypothesis $H_i$ over the alternative. For instance, a Bayes factor of 10 means the alternative hypothesis $H_1$ is 10 times more likely given this observed data compared to the null hypothesis. Typical decision threshold are kept at 10 or 30 (1/10 or 1/30 respectively). The direct comparison between the two hypothesis obviates the need for multiple testing correction, as both hypothesis are now compared like-for-like.

However, there are two drawbacks to the Bayesian approach. First, evaluating the evidence for both the null and alternative hypotheses can be computationally demanding. Second, we have given up control of our type I error. Fortunately, for many hypothesis tests, the evaluation of the evidence is not required. Additionally, whereas a type I error cannot be controlled explicitly, it can be inferred through simulations. However, for the sequential hypothesis, Bayesian models need to be constructed for every step of the sequence of each simulation, which we will leave for future work.

In case a Bayes Factor is used for a hypothesis test, where the null model is a special case of the alternative model, the Bayes Factor simplifies substantially. We assume the following hypothesis for a particular model parameter $\lambda$ of interest: $H_0 : \lambda = \lambda_0$, and $H_1 : \lambda \neq \lambda_0$ ($\lambda$ is anything). Here $H_0$ is a special case of $H_1$ and the hypotheses are *nested*. (2) Assuming equal priors for the hypotheses, we can write

$$BF_{01} = \frac{p(\boldsymbol{D}|H_1)}{p(\boldsymbol{D}|H_0)} = \frac{p(\boldsymbol{D}|H_1)}{p(\boldsymbol{D}|\lambda = \lambda_0, H_1)}.$$

Applying Bayes' rule to the denominator $p(\boldsymbol{D}|\lambda = \lambda_0, H_1) = \frac{p(\boldsymbol{D}|H_1)p(\lambda=\lambda_0|\boldsymbol{D},H_1)}{p(\lambda=\lambda_0|H_1)}$, we arrive at

$$BF_{01} = \frac{p(\lambda = \lambda_0|H_1)}{p(\lambda = \lambda_0|\boldsymbol{D}, H_1)}, (3)$$

where $p(\lambda = \lambda_0|H_1)$ is simply the prior probability of $\lambda = \lambda_0$ and $p(\lambda = \lambda_0|\boldsymbol{D}, H_1)$ is simply the probability of $\lambda = \lambda_0$ under the posterior for $\lambda$. The evaluation of the prior is cheap, the posterior is usually the result of evaluating the Bayesian model.

More generally, the *nested* hypothesis can be used for disjoint hypothesis test of the form: $H_0 : \lambda < \lambda_0$, and $H_1 : \lambda \geq \lambda_0$ via the use of an encompassing hypothesis $H_e$ : $\lambda$ is anything, in which both hypotheses are nested. (3) Writing this out and using Eqn. 3:

$$BF_{01} = \frac{p(\boldsymbol{D}|H_1, \lambda \geq \lambda_0)}{p(\boldsymbol{D}|H_{0,} \lambda < \lambda_0)} = \frac{p(\boldsymbol{D}|H_e, \lambda \geq \lambda_0)}{p(\boldsymbol{D}|H_e)} \cdot \frac{p(\boldsymbol{D}|H_e)}{p(\boldsymbol{D}|H_{e,} \lambda < \lambda_0)}$$

$$BF_{01} = \frac{p(\lambda < \lambda_0|H_e)p(\lambda \geq \lambda_0|\boldsymbol{D}, H_e)}{p(\lambda < \lambda_0|\boldsymbol{D}, H_e)p(\lambda \geq \lambda_0|H_e)}.$$

Note that the encompassing hypothesis is typically assumed through the Bayesian modelling.

# S8. Sensitivity analysis priors

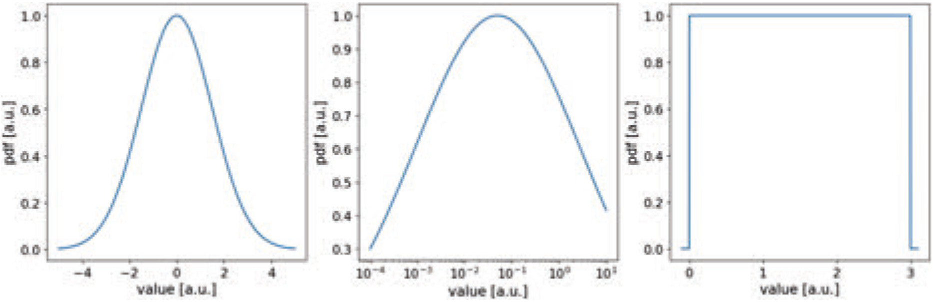


**Figure S10.** Less informative priors used for parameters in modeling (c.f. Figure S1). A: β (Normal $\mu = 0, \sigma = 1.5$), B: b (LogNormal $\mu = 13, \sigma = 4$), and C: k (Uniform, [0, 3])

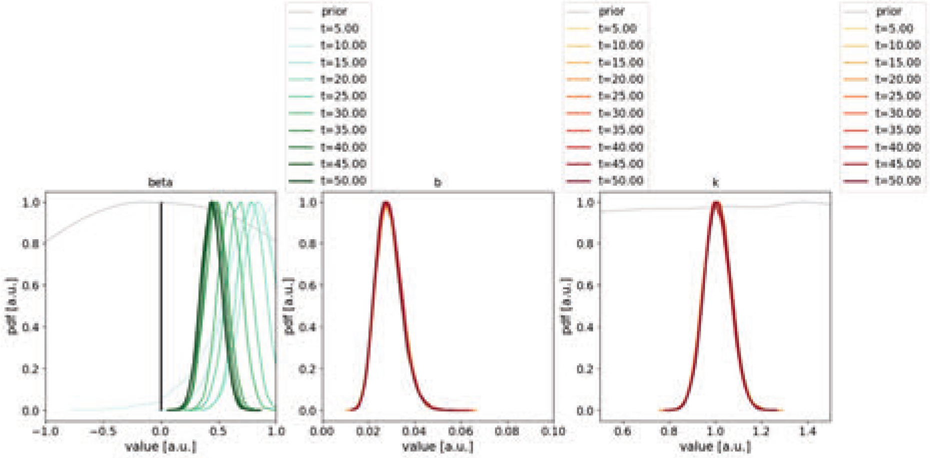


**Figure S11.** Posteriors with less informative priors on dataset 1 (c.f. Figure S6). A: β, B: b, and C: k. Since the trial cohort is included fully at $T = 0$, the posteriors for the baseline parameters $b$, and $k$ are thus virtually unaffected by the change in prior.

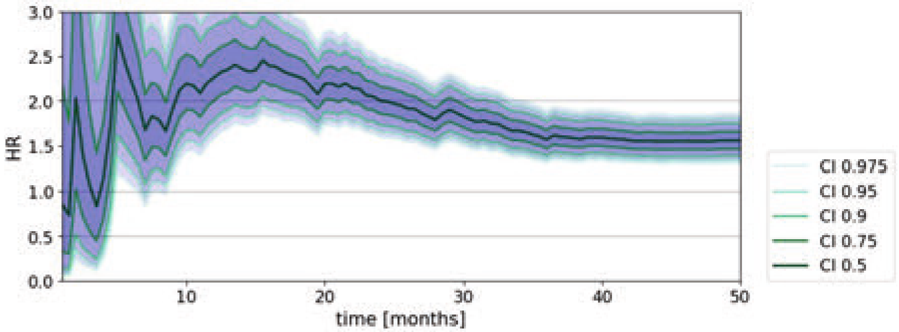


**Figure S12.** Credible intervals for Dataset 1, given the less informative priors from Figure S10 (c.f. Figure 2A in main text). Initial, in the first 5 months, the effect of the substantially less informing prior is substantial. From about 10 months this effect of this prior is diminished, as more samples start contributing to the posterior of the model.

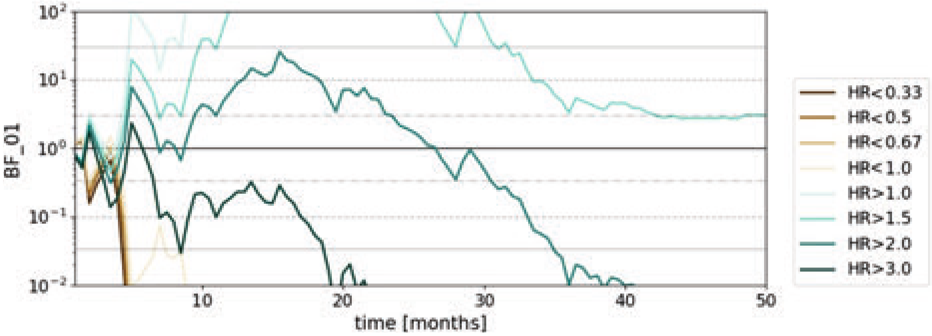


**Figure S13.** Bayesian Factors for Dataset 1, given the less informative priors from Figure S10 (c.f. Figure 2B in main text). Because of the initial inclusion of a large number of patients in this dataset, the decision results become rapidly (within about 5 months) similar to the decisions derived from the more informative and realistic prior.

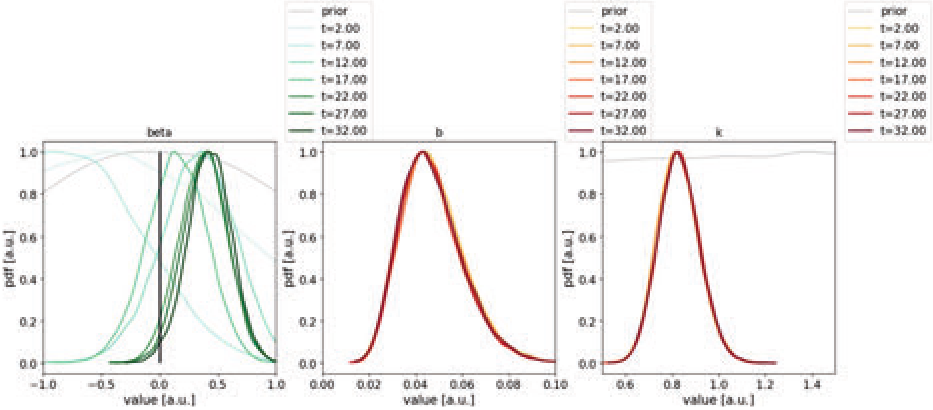


**Figure S14.** Posteriors with less informative priors on Dataset 2 (c.f. Fig. S8). A: β, B: b, and C: k. Since the trial cohort is included fully at $T = 0$, the posteriors for the baseline parameters $b$, and $k$ are thus virtually unaffected by the change in prior.

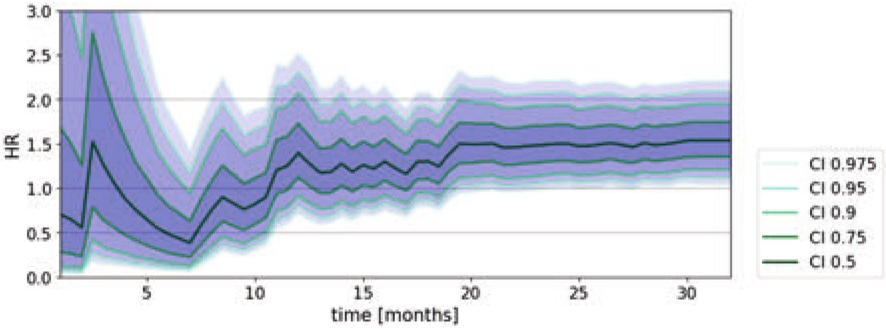


**Figure S15.** Credible intervals for Dataset 2, given the less informative priors from Figure S10 (c.f. Figure S20A). Initially, in the first 5 months, the effect of the substantially less informing prior is substantial. From about 10 months this effect of this prior is diminished, as more samples start contributing to the posterior of the model.

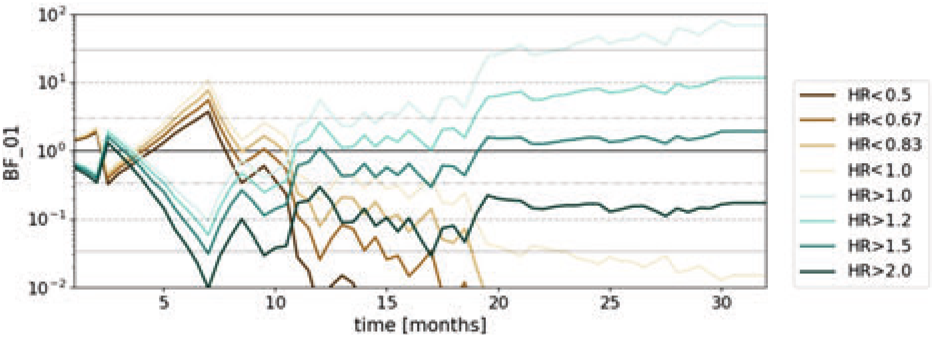


**Figure S16.** Bayesian Factors for Dataset 2, given the less informative priors from Figure S10 (c.f. Figure S20B). Initial, in the first 5 months, the effect of the substantially less informing prior is substantial, pushing some decisions towards the cut-off thresholds. This stems from the wide prior over β.

## S9. Posteriors using Gompertz baseline hazard function

To verify the independence of the results from the choice of baseline hazard (Weibull), we also ran our model against a Gompertz baseline hazard function, parameterized as

$$h_0(t|\boldsymbol{m}) = h(t|b,\eta) = b\eta e^{bt}.$$

Figure S17 and S18 show the resulting credible intervals around $\beta$ of the modelling using the Gompertz baseline hazard function described above (with the priors for $b$ and $\beta$ as described in SI section 1 and the prior for $\eta$ taken from a uniform distribution on support $[1,5]$) applied to datasets 1 and 2. When comparing these figures to Figure 2A in the main text and Figure S20A, no visually meaningful differences are observed between the modelling using a Gompertz or Weibull baseline hazard function.

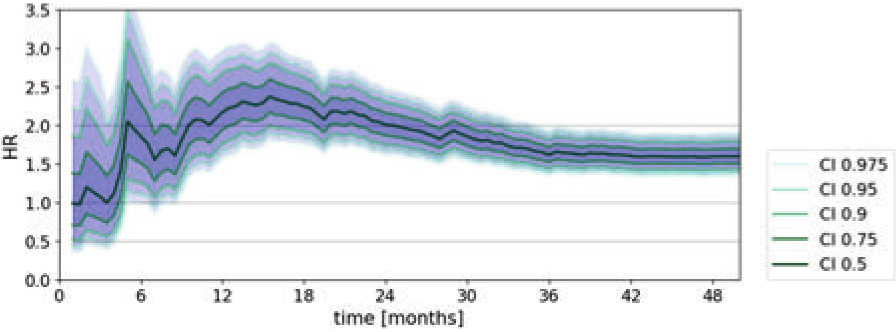


**Figure S17.** Credible intervals for Dataset 1 using the Gompertz baseline hazard (c.f. Figure 2A in the main text). No visual difference is observed compared to the modelling using the Weibull baseline hazard.

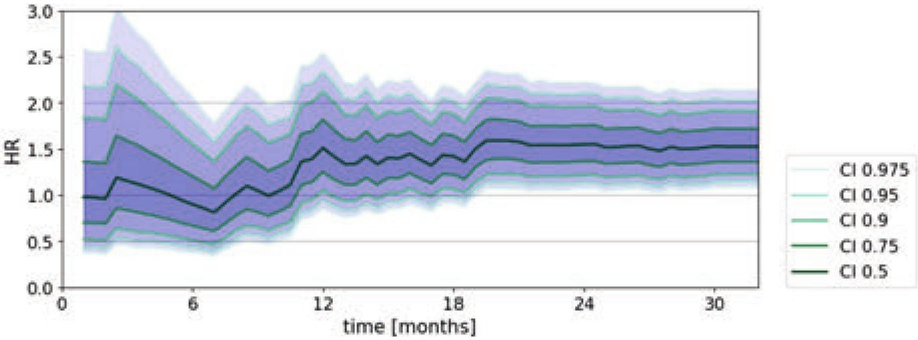


**Figure S18.** Credible intervals for Dataset 2 using the Gompertz baseline hazard (c.f. Figure S20A). No visual difference is observed compared to the modelling using the Weibull baseline hazard.

## S10. Distribution of patient and disease characteristics over time in clinical practice

**Table S3.** Patient and disease characteristics over time for patients treated with first line pembrolizumab plus chemotherapy in clinical practice from dataset 1.

| Time (months) | Total population N | Age ≥ 75 year n (%**) | Males n (%**) | Squamous n (%**) | PD-L1 <1% n (%) | ECOG PS ≥ 2 n (%**) | Brain metastasis yes n (%**) |
|---|---|---|---|---|---|---|---|
| 5 | 53 | 4 (7) | 29 (53) | 0 | 23 (42) | 7 (13) | 12 (22) |
| 10 | 141 | 12 (8) | 68 (48) | 0 | 42 (29) | 14 (10) | 32 (22) |
| 15 | 212 | 22 (10) | 108 (50) | 0 | 118 (55) | 17 (8) | 44 (20) |
| 20 | 265 | 28 (10) | 133 (49) | 0 | 145 (54) | 21 (8) | 59 (22) |
| 25 | 335 | 35 (10) | 166 (49) | 0 | 185 (55) | 27 (8) | 70 (21) |
| 30 | 415 | 49 (12) | 200 (48) | 0 | 224 (53) | 35 (8) | 79 (19) |
| 35 | 484 | 62 (13) | 237 (48) | 0 | 257 (52) | 44 (9) | 93 (19) |
| 40 | 512 | 67 (13) | 244 (48) | 0 | 269 (53) | 47 (9) | 96 (19) |
| 45 | 512 | 67 (13) | 244 (48) | 0 | 269 (53) | 47 (9) | 96 (19) |
| 50 | 512 | 67 (13) | 244 (48) | 0 | 269 (53) | 47 (9) | 96 (19) |

* The total number of patients enrolled in the study since the start of the real-world cohort follow-up (T0), which corresponds to the date when the first patient *initiated* treatment with pembrolizumab.

** % calculated by dividing n / N

**Table S4.** Patient and disease characteristics over time for patients treated with first line pembrolizumab in clinical practice from dataset 2.

| Time (months) | Total population N | Age ≥ 75 year n (%**) | Males n (%**) | Squamous n (%**) | PD-L1 <1% n (%) | ECOG PS ≥ 1$ n (%**) | Brain metastasis yes n (%**) |
|---|---|---|---|---|---|---|---|
| 5 | 19 | 2 (11) | 11 (58) | 3 (33) | 0 (0) | 14 (74) | 6 (32) |
| 10 | 46 | 5 (11) | 27 (59) | 9 (20) | 0 (0) | 31 (67) | 11 (24) |
| 15 | 73 | 9 (12) | 42 (57) | 10 (14) | 0 (0) | 48 (65) | 15 (20) |
| 20 | 82 | 10 (12) | 45 (55) | 11 (13) | 0 (0) | 52 (63) | 15 (18) |
| 25 | 82 | 10 (12) | 45 (54) | 11 (13) | 0 (0) | 53 (64) | 16 (19) |
| 30 | 83 | 10 (12) | 45 (54) | 11 (13) | 0 (0) | 53 (64) | 16 (19) |

**Abbreviations**: PD-L1, Programmed death-ligand 1; ECOG PS, Eastern Cooperative Oncology Group Performance status.

*The total number of patients enrolled in the study since the start of the clinical practice cohort follow-up (T0), which corresponds to the date when the first patient initiated treatment with pembrolizumab.

** % calculated by dividing n / N

$ Since only two patients have an ECOG PS of 2 or higher in this dataset, we choose ECOG PS ≥ 1 (yes or no) as variable of interest.

# S11. Hazard ratios and 95% credible intervals at subsequent time points for dataset 1 and 2

**Table S5.** Median HR with 95% credible intervals, number of patients (n) for dataset 1. The HR represent overall survival difference for patients treated with pembrolizumab plus chemotherapy in clinical practice versus KEYNOTE-189 RCT

| Time (months) | Median HR (95% credible intervals) | Patients (n) from clinical practice data | | | |
|---|---|---|---|---|---|
| | | Total* | at risk | event | censored |
| 6 | 1.78 (1.06 – 3.22) | 72 | 61 | 10 | 1 |
| 12 | 1.98 (1.40 – 2.66) | 171 | 118 | 46 | 7 |
| 18 | 2.18 (1.73 – 2.73) | 239 | 130 | 97 | 12 |
| 24 | 1.91 (1.60 – 2.27) | 320 | 165 | 139 | 16 |
| 30 | 1.82 (1.58 – 2.16) | 415 | 202 | 191 | 22 |
| 36 | 1.78 (1.52 – 2.10) | 499 | 234 | 228 | 37 |
| 42 | 1.60 (1.38 - 1.86) | 512 | 227 | 274 | 194 |
| 48 | 1.58 (1.36 – 1.82) | 512 | 1 | 276 | 235 |
| 50 | 1.54 (1.33 – 1.78) | 512 | 0 | 277 | 235 |

*Abbreviations: Hazard Rate, HR. * The total number of patients enrolled in the study since the start of the clinical practice cohort follow-up, which corresponds to the date when the first patient initiated treatment with pembrolizumab pus chemotherapy.*

**Table S6.** Median HR with 95% credible intervals, number of patients (n) for dataset 2. The HR represent overall survival difference for patients treated with pembrolizumab in clinical practice versus KEYNOTE-024 RCT.

| Time (months) | Median HR (95% credible interval) | Patients (n) from clinical practice data | | | |
|---|---|---|---|---|---|
| | | Total* | at risk | event | censored |
| 6 | 0.96 (0.46 – 1.95 | 23 | 22 | 1 | 0 |
| 12 | 1.00 (0.55 – 1.70) | 57 | 45 | 12 | 0 |
| 18 | 1.19 (0.78 – 1.80) | 80 | 56 | 24 | 0 |
| 24 | 1.45 (1.07 – 1.96) | 82 | 43 | 39 | 1 |
| 30 | 1.49 (1.11 – 1.98) | 83 | 0 | 48 | 35 |

*Abbreviations: Hazard Rate, HR. * The total number of patients enrolled in the study since the start of the clinical practice cohort follow-up, which corresponds to the date when the first patient initiated treatment with pembrolizumab*

## S12. Results dataset 2: survival outcomes associated with pembrolizumab monotherapy

### Kaplan-Meier curves

The Kaplan-Meier curves of overall survival outcomes associated with pembrolizumab monotherapy for the RCT cohort and the complete clinical practice dataset (i.e., comprising of all patients and complete follow-up) are presented in Figure S19.

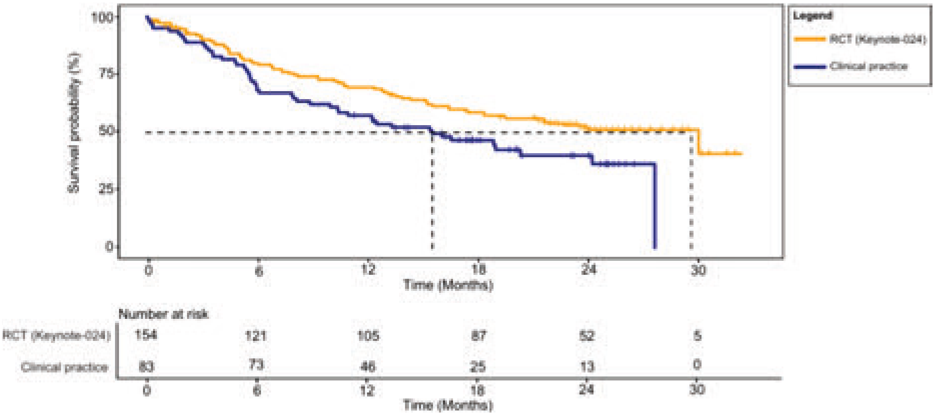


**Figure S19.** Kaplan-Meier curves of overall survival outcomes for patients treated with pembrolizumab monotherapy in clinical practice versus the KEYNOTE-024 RCT.

### Model diagnostics

The MCMC procedure demonstrated good convergence and resolution, as detailed in Table S3 and Figure S4. Figure S5 shows that the predictive prior checks demonstrated good alignment between the prior and the observed data, indicating that the chosen priors are appropriate. Figure S9A-E provide a visualization of the model fit at subsequent time intervals, through posterior predictive checks. Although there were some small deviations, the overall fit was good.

### Posterior distribution

Figure S20A shows the HR estimates for each consecutive model with credible intervals from 50% to 97.5%. Initially, in the first few months when limited data from clinical practice was included (i.e., a small number of patients and short follow-up), the credible intervals indicated large uncertainty and HR estimates fluctuated. As more data accumulated (i.e., more patients and longer follow-up), uncertainty decreased and the HR distribution shifted in its entirety towards HR > 1, indicating with high certainty a worse survival for patients treated in clinical practice compared to those treated in the RCT. After about 20 months of data accumulation, the HR distribution reached a stable value of approximately 1.50. When all data were included, the HR was 1.49 (95% credible interval 1.11 – 1.96) (Table S4).

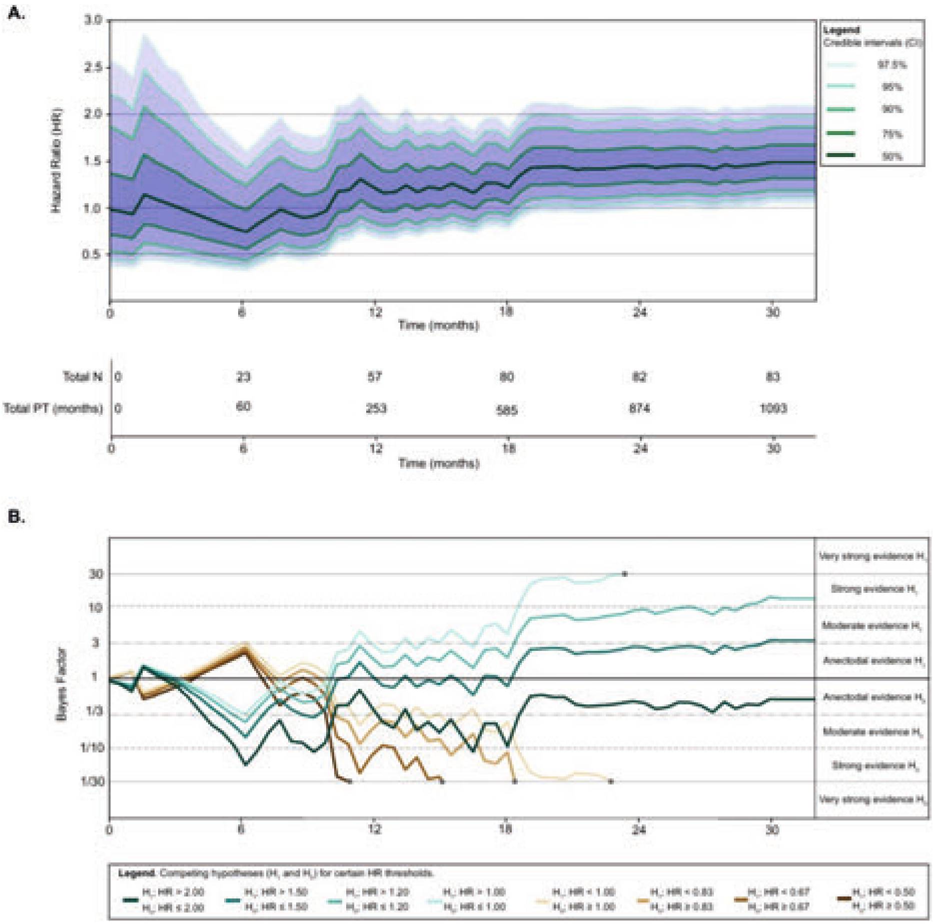


**Figure S20.** Hazard Ratio (HR) represents the hazard of a survival event occurring in the clinical practice cohort divided by that in the RCT cohort for patients treated with pembrolizumab monotherapy A: HR distribution over time with 50%, 75%, 95% and 97.5% credible intervals. B: The Bayes factor for competing hypotheses testing different HR thresholds for each model update. The grey circle indicates the point at which there is very strong evidence for either $H_1$ or $H_0$ for every further updated model.

## Bayesian hypothesis testing

Figure S20B shows the of the Bayes Factor over time for different competing hypotheses, evaluating whether the observed HR falls above certain thresholds (ranging from HR >1.00 to >2.00) or below certain thresholds (ranging from HR <1.00 to <0.50). In the first 10 months data accumulation, there was a lack of evidence to support a HR > 1.00, indicating worse survival outcomes in clinical practice than in the RCT (light green curve), or a HR < 1.00, indicating better survival outcomes (light orange curve). After 10 months of data accumulation, the strength of the evidence for an HR > 1.00 increased. Before the end of data accumulation, there was very strong evidence for an HR > 1.0 and strong evidence for a HR >1.2.

### Sensitivity analyses

The prior sensitivity analysis, detailed in Figure S15 and S16 demonstrated that variations in the range of priors did not impact the results, confirming the robustness of the chosen prior.

## S13. References

1. Mackay, D.J.C. *Information theory, inference and learning algorithms.* 6th edn. (Cambrigde university press, UK, 2003)
2. Wagenmakers, EJ. *et al*. Bayesian hypothesis testing for psychologists: A tutorial on the Savage-Dickey method. *Cogn Psychol.* **60**, 158-189 (2010).
3. Klugkist, I and Kato, B and Hoijtink, H. Bayesian model selection using encompassing priors. *Stat Neerlan.* **59**, 57-69.
4. Petri, H, Urquhart J. Channeling bias in the interpretation of drug effects. *Stat med.* **10.** 577-581.

# CHAPTER 4

# General discussion

Treatment effects observed in daily clinical practice often differ from randomized controlled trials (RCTs) because of more heterogenous populations, where patients are usually older, have more comorbidities, are in worse performance states, and receive less consistent treatment schedules and follow-up procedures. These differences highlight the need for real-world data (RWD) to complement traditional pre-approval clinical trials by providing insights into how treatments perform in broader, more diverse populations and in everyday healthcare settings. [1-3] In addition, RWD gives the opportunity to compare treatments in the absence of head-to-head RCTs. Generating these insights promptly and with high validity is essential for decision-making regarding the appropriate use of medicines, as done by a variety of relevant parties, including regulatory authorities, policymakers, and payers. By "prompt evidence generation," we refer to the ability to obtain insights into survival outcomes and tolerability as soon as possible after introduction of novel therapies in clinical practice.

This thesis focused on real-world evidence (RWE) to support patients and clinicians in the context of clinical decision-making. To do so, the aim of this thesis was to complement data on survival outcomes reported in RCTs of patients with advanced NSCLC with data from patients treated with immunotherapy in clinical practice, while also evaluating methodological challenges and advancements for prompt evidence generation.

More specifically, the first part of this thesis, **Chapter 2**, assessed real-world outcomes of patients with NSCLC treated with immunotherapy. In **Chapter 2.1**, the real-world effectiveness of durvalumab after concurrent or sequential chemoradiotherapy was evaluated in patients with stage III NSCLC. Durvalumab treatment in clinical practice was associated with improved progression-free survival (PFS) and overall survival (OS) compared with historical controls treated with CRT alone. The median PFS observed in clinical practice was notably longer than that reported in the PACIFIC trial, while OS outcomes were similar between the two settings. In **Chapter 2.2**, the real-world effectiveness of first line pembrolizumab plus chemotherapy was compared with the efficacy reported in the corresponding RCTs (KEYNOTE-189 and KEYNOTE-407). Patients with nonsquamous PD-L1-negative tumors who were treated in clinical practice experienced considerably shorter OS than those in the corresponding trial. No statistically significant difference was observed among subgroups of patients with PD-L1-expressing tumors. In **Chapter 2.3**, treatment response and PFS were compared between patients receiving first line nivolumab–ipilimumab plus chemotherapy and those receiving pembrolizumab plus chemotherapy in clinical practice. The dual immunotherapy regimen demonstrated a numerically higher response rate and similar PFS compared with pembrolizumab plus chemotherapy. In **Chapter 2.4**, the potential effect of concomitant medication on the effectiveness of immunotherapy was investigated. To address potential confounding by indication, the immunotherapy cohort was matched to a chemotherapy cohort. Patients treated with antibiotics or opioids showed worse survival outcomes in both cohorts, suggesting that the association may reflect underlying patient characteristics rather than a specific immunomodulating effect. Altogether, these studies provide valuable insights into real-world outcomes of immunotherapy in patients with advanced NSCLC, offering evidence that clinicians can incorporate into their decision-making.

The second part of this thesis, **Chapter 3**, addressed methodological challenges and advancements for prompt evidence generation. In **Chapter 3.1**, a scoping review showed that observational studies often lack clear reporting of PFS definitions and measurements, and that, when reported, these definitions frequently differ from those used in RCTs. In **Chapter 3.2**, the use of text-mining techniques was explored to extract structured information from unstructured clinical notes, demonstrating that this approach can support more standardized and efficient data capture. In **Chapter 3.3**, a Bayesian survival model was developed to compare accumulating time-to-event data from clinical practice with fixed RCT data. This approach showed promise for reducing the time needed to assess whether differences in OS between clinical practice and RCT populations exceed predefined thresholds.

The work presented in this thesis provides valuable insights into the real-world outcomes of patients with advanced NSCLC treated with immunotherapy, while also demonstrating how evidence regarding these outcomes can be generated more promptly. At the same time, the findings and their context reveal recurring challenges, as PFS results may vary and not fully capture what matters most to patients, while inconsistencies in reporting and measurement can undermine validity, and conventional approaches to data collection and analysis remain too slow to keep pace with a rapidly evolving treatment landscape.

These observations highlight the need for *meaningful*, and *valid outcomes*, and *prompt evidence generation.* In the following paragraphs, these three aspects are further elaborated, placed in the context of the studies described above, and accompanied by recommendations for future directions.

## Meaningful outcomes

In this thesis, OS and PFS were the main clinical outcomes studied. These outcomes represent the primary endpoints most frequently assessed in pre-approval studies. However, it is questionable how meaningful these outcomes are for the individual patient. Traditionally, OS has been considered the gold standard in oncology research, as longer survival is generally considered to reflect treatment benefit. Therefore, OS will remain an important outcome. [4] However, OS does not capture the effect of a treatment on the quality of survival, which patients often value as much as, or even more than, survival itself. [5] Another limitation of OS is that it is a composite of all treatments provided to a patient (e.g., post-progression treatments following first-line treatment), which limits the assessment of the contribution of a specific treatment. [6] For this reason, PFS is commonly used because it reflects the effect of a single treatment. However, the correlation between PFS and OS, as well as between PFS and quality of life, remains uncertain in RCTs and is equally questionable in clinical practice. [7-9] The uncertainty regarding the correlation between PFS and OS in clinical practice was illustrated in **Chapter 2.3**, where patients treated with durvalumab in clinical practice had a notably longer PFS than those treated in the reference trial (27.1 vs. 16.8 months), while OS outcomes were comparable.

The ongoing debate about clinical outcomes in cancer research revealed that patients often prioritize health-related quality-of-life outcomes, such as physical functioning, symptom control, emotional well-being, and the ability to maintain daily activities and social connections, over survival itself. [10] These aspects are best captured through patient-reported outcome measures (PROMs), which directly assess patient outcomes without interpretation by clinicians or others. [11]. Therefore, clinical decision-making should be guided not only by survival data but also by PROM data, ensuring that treatment decisions reflect what matters most to patients. [12-13] However, despite the increasing recognition of the value of PROMs, they are still not routinely used as outcomes in RCTs, and even less so in observational studies [14]. This limited uptake is mainly due to concerns about their reliability, both methodological and practical. Recent developments, however, are actively addressing these challenges. Methodological challenges are being addressed through the development of international standards such as SISAQOL-IMI, which harmonize the reporting, interpretation, and analysis of PROM data. [15] Practical challenges are being addressed through innovations such as decentralized and hybrid trials with remote data-collection platforms, as well as projects like Trials@Home. [16] Additionally, national initiatives such as the Dutch Thoracic Oncology Cohort (DuTOC) are establishing an infrastructure for prospectively capturing data, including PROM data from patients with lung cancer in clinical practice. These patients will periodically receive validated digital PROMs to enable longitudinal data collection. [17] With the first patients enrolled in August 2025, this initiative is still in its early phase but already holds considerable promise. Similar efforts in other cancer types have demonstrated the feasibility of such large-scale, prospective PROM data collection. For example, in colorectal cancer, the Prospective Dutch Colorectal Cancer cohort has shown that this approach is achievable in clinical practice. [18] Building on these experiences, extending similar approaches to lung cancer would represent an important step forward. At the same time, the automated integration of PROMs into

EHRs remains challenging, and further investments in digital infrastructure are needed to make PROM assessment easier for patients, ideally through passive capture with direct integration into EHR systems. These priorities are in line with broader initiatives such as the European Health Data Space and Health-RI in the Netherlands, which aim to create interoperable infrastructures for health data and thereby support the integration and use of PROM data in both research and routine care. [19, 20]

Despite the challenges associated with PROM data, their continued development and use are essential. In this thesis, PROM data were not collected. Their inclusion in **Chapter 2.3** would have been particularly valuable, as OS data were not mature at the time of data analysis, and PROMs data could have provided complementary insights into how nivolumab–ipilimumab plus chemotherapy and pembrolizumab plus chemotherapy affect quality of life in clinical practice. This could be different because of the additional CTLA-4 target and different number of chemotherapy cycles. Until PROMs are more widely implemented, outcomes such as OS and PFS remain the most used clinical endpoints, although both present considerable challenges in reflecting treatment effectiveness in real-world settings. These challenges and other potential biases will be discussed in the next paragraph.

## Valid outcomes

In this thesis, we used observational cohort designs to investigate the effectiveness of immunotherapy in patients with advanced NSCLC, either by comparing two or more clinical practice cohorts (i.e. within study comparison), or by comparing a clinical practice cohort with an RCT cohort (i.e. between studies comparison). While observational studies offer valuable insights, they are inherently more vulnerable to bias than RCTs since they lack randomization and blinding. [21] The main biases in observational studies are selection bias, information bias and confounding bias. The extent to which these biases impact the validity of a study depends on the specific study design and research question. Here, these types of bias are discussed in the light of the studies in this thesis. Also, the specific strategies used to mitigate their impact are outlined. Finally, a more general approach that could be used to mitigate bias in such observational studies is outlined.

First, information bias, often referred to as non-random misclassification, occurs when study variables (covariates or outcomes) are systematically measured, reported or classified incorrectly. [22] Regarding the primary outcomes used in this thesis, OS is generally not affected by information bias because death is clearly defined and objectively captured, whereas PFS is more prone to such bias, particularly in real-world studies, where assessments are less standardized than in RCTs. PFS can be influenced by several factors, such as the presence of clinical symptoms, the timing of radiological assessments, the criteria used to interpret scans, and the individuals performing these interpretations. [23, 24] These factors vary not only across different clinical practice settings but also between clinical practice and the highly controlled RCT environment (see **Chapter 3.1**). This variability makes it particularly challenging (if not impossible) to accurately compare PFS outcomes between observational and RCT cohorts (**Chapters 2.2 and 2.3**). This might also explain the unexpected finding that PFS for patients treated with durvalumab in clinical practice was longer than for patients treated with durvalumab in the RCT (**Chapter 2.1**). These observations raised concerns about both the reliability and accuracy of PFS measurements in real-world settings. Therefore, when PFS is studied, appropriate strategies should be taken to reduce bias. First, factors influencing PFS measurement should be aligned as much as possible. If the timing and frequency of assessments cannot be harmonized, evaluation-time bias can be minimized by applying interval-censoring rules and using appropriate statistical methods, as recommended by Panageas et al. and Degue et al. [25, 26] Alternatively, more objective early outcome measures, such as time to treatment discontinuation or time to next treatment, may be considered. Unlike PFS, these outcomes are less prone to information bias, as they rely on medication orders which are often objectively recorded. [27] However, similar to PFS, no correlation with OS or quality of life has been demonstrated, raising questions about their value as meaningful outcomes for patients.

Second, selection bias occurs when the likelihood of being included in either the treatment or the comparator group of a comparative effectiveness study is related to factors that also influence the outcome, thereby threatening the internal validity of the study. [28-29] The comparative effectiveness study in **Chapter 2.1** provides an example where the internal validity was threatened by immortal time bias, often referred to as a form of selection bias. Immortal time bias arises when there is a period during follow-up in which a time-to-event outcome (such as death or disease progression) cannot occur, yet, this time is incorrectly classified as part of the time at risk. [30] One of the study's objectives was to compare the survival outcomes of patients who received CRT followed by adjuvant durvalumab with those of a historical control group who received CRT alone. In the durvalumab group, patients could only receive the treatment if they survived the period between the start of CRT and the initiation of durvalumab, which is referred to as "immortal time." In contrast, patients in the historical control group did not experience this "immortal time," potentially leading to an overestimation of survival in the durvalumab cohort. To minimize this bias, we excluded all patients in the historical control cohort who had progressed, died, or were censored before an imputed index date.

Third, confounding bias occurs when a variable is associated with both the exposure and the outcome, distorting the true association between them. [31] A commonly described type of confounding in real-world studies is confounding by indication, which occurs when treatment allocations are influenced by disease or patient characteristics that are also associated with treatment outcomes. [32, 33] Confounding by indication or 'treatment selection bias' is closely related to selection bias. [34] In **Chapter 2.4**, this thesis provides an example where we did not aim to eliminate confounding by indication but instead used this principle to assess the impact of comedication influencing the gut microbiome on the effectiveness of immunotherapy to evaluate potential effect modification. It was hypothesized that certain comedications, such as antibiotics or proton pump inhibitors, may alter the composition of the gut microbiome in ways that diminish the immune response, thereby reducing the effectiveness of immunotherapy. We used a matched cohort design of a chemotherapy and immunotherapy cohort, to evaluate the relation between the use of gut microbiome-influencing comedication and survival outcomes within each cohort. In both cohorts, the use of antibiotics and other medication was similarly associated with worse outcomes. This suggests that the association was likely a consequence of confounding rather than comedication disturbing the composition of the microbiome and thus no effect modification. In **Chapter 2.2**, this thesis presents an example of how bias arising from treatment allocations based on PD-L1 expression can be minimized. At the time of the study, guidelines recommended pembrolizumab plus chemotherapy for patients with no (PD-L1 <1%) or intermediate (PD-L1 1%-49%) PD-L1 expression. For patients with PD-L1 > 50%, pembrolizumab monotherapy was recommended unless a rapid tumor response was required, in which case pembrolizumab plus chemotherapy could be considered. As a result, a significantly lower proportion of patients with PD-L1 >50% were exposed to pembrolizumab plus chemotherapy in clinical practice compared to RCTs (11.5% versus 31.2%). Given that PD-L1 >50% is associated with better survival outcomes than lower PD-L1 expression and that treatment choices depend on PD-L1 status, differences in survival outcomes between clinical practice and RCTs were assessed by PD-L1 stratum to prevent confounding by indication.

The examples above highlight some of the most common biases and the specific strategies used to mitigate them. However, each study comes with its own methodological challenges, which must be addressed to generate valid RWE for clinical decision-making. A relatively recent methodological development that can help overcome some of these challenges is the target trial emulation (TTE) approach. [35] In this approach, observational studies are explicitly designed to emulate the key elements of a hypothetical RCT, including eligibility criteria, treatment strategies, follow-up, and outcome definitions. Initiatives such as REPEAT and DUPLICATE have demonstrated that when observational studies are carefully designed using a TTE approach, they can yield results that closely align with those from a corresponding RCT. [36, 37] The robustness of this approach has also been demonstrated in metastatic NSCLC, where Polito et al.

(2024) emulated a trial comparing OS of patients treated with first-line chemotherapy in clinical practice with those treated in RCTs, resulting in comparable survival estimates. [38] That emulations of conducted clinical trials yield similar findings adds to the trust in this methodology to also study comparisons that are not investigated in head-to-head clinical trials. [39] The comparative effectiveness study presented in **Chapter 2.3** that compared nivolumab–ipilimumab plus chemotherapy with pembrolizumab plus chemotherapy in advanced NSCLC patients with PD-L1 <1%, already incorporated a TTE principle by beforehand defining response criteria to be handled identically per treatment arm. While the timing of follow-up assessments is difficult to standardize in clinical practice, applying consistent response minimized bias.

### Prompt evidence generation

As discussed above, valid study designs and meaningful outcomes are essential for informing clinical decision-making. With new treatment options rapidly emerging, it becomes even more important to inform patients and clinicians promptly about their effectiveness in clinical practice. However, prompt real-world evidence generation is often hindered, as most studies (including those in this thesis) rely on EHR data from a limited number of centers, resulting in relatively small patient populations. Consequently, cohort recruitment is slow, delaying the availability of reliable evidence. Collaborations with multiple centers could help to overcome this. However, manual EHR data collection across multiple centers is also time-consuming and prone to error, as data items are often recorded and interpreted differently across centers. Therefore, data standardization is needed to ensure that data collected across centers are comparable, reliable, and can be combined for robust analyses. [40] The International Consortium for Health Outcomes Measurement (ICHOM) promotes the standardization of outcomes and data structures across a wide range of diseases worldwide. [41] At the national level, the Netherlands Comprehensive Cancer Organisation (IKNL) supports such efforts in oncology, including in the field of lung cancer. [42] Although promising, the uptake of these data standards in clinical practice has been slow, probably due to the complexity of implementation. While such initiatives define which outcomes should be measured and how they can be collected, Common Data Models (CDMs) further support data harmonization by organizing information into a standard structure with shared terminology. [43] Although CDMs provide a common framework for routinely collected data, they often lack the flexibility to incorporate new or study-specific variables, particularly those documented in unstructured text. [44] This limitation is especially relevant for clinical outcomes such as time to disease progression or symptom burden, which are rarely noted in structured formats. In such cases, text mining offers a scalable solution by converting unstructured data into structured formats, thereby complementing CDMs and expanding their utility. As demonstrated in **Chapter 3.2**, we successfully applied text mining to extract information on disease progression from unstructured medical and radiological notes in multiple centers. At the same time, new opportunities are emerging to improve data capture at the source. Digital health technologies, such as medical wearables, can reduce information bias by enabling more complete and objective outcome collection, thereby limiting the variability introduced through voluntary manual documentation. [45] Combined with text mining and CDMs, these technologies can broaden the scope and quality of real-world data, ultimately providing more reliable evidence for patients and clinicians. National initiatives also play a key role in advancing real-world evidence generation. As previously highlighted, DuTOC provides such an example by creating a prospective infrastructure for clinical data, PROMs, and biomaterials, thereby supporting timely and reliable evidence in lung cancer.

Although these strategies accelerate access to larger and more reliable datasets, valid conclusions about treatment effectiveness still require time. In most cohort studies, analyses are typically tied to fixed datasets with predefined follow-up, resulting in conclusions that often arrive late. These delayed conclusions are especially undesirable in the rapidly evolving landscape of lung cancer treatment. This limitation could be addressed by continuously monitoring and comparing the effectiveness of new treatments in real time, either relative to the pivotal trial or between different treatment strategies. Real-time monitoring would allow early identification of potential efficacy-effectiveness gaps and enable timely adjustments to treatment strategies. Stratifying these real-time analyses by patient subgroups and clinical settings would enable the generation of prompt insights to support individualized treatment decisions. If this technique had been applied in the study on the effectiveness of pembrolizumab plus chemotherapy (**Chapter 2.2**), it would likely have identified much earlier that patients with PD-L1 <1% treated in clinical practice had significantly worse OS than those treated in the RCT, as follow-up could have started from the first patient treated in 2019. In this thesis, **Chapter 3.3** presents a first step towards real-time monitoring by applying a Bayesian survival approach that compares survival outcomes from an accumulating clinical practice cohort with survival outcomes from a fixed RCT cohort. This type of real-time monitoring, combined with data standardization and harmonization, is expected to accelerate the availability of relevant evidence and help patients and clinicians make better-informed treatment decisions soon after the adoption of new treatments.

## Recommendations for research

Future research should include more outcomes that are truly meaningful for patients with advanced NSCLC. This requires systematic collection and use of PROM data, supported by initiatives such as DuTOC. In addition, national initiatives should aim for automatic integration of PROM into EHRs, ideally through passive digital capture (e.g. through wearables or patient apps). At the same time, methodological research should focus on strengthening the validity of observational studies. Methods such as TTE should be investigated for their potential to reduce bias and enable real-world treatment comparisons when head-to-head trials are lacking. Ultimately, research strategies should focus on promptly generating evidence, which can be achieved if variables and outcomes are harmonized and standardized as much as possible. Advanced techniques, such as text-mining, can be used to extract and structure variables and outcomes and Bayesian survival modelling can be used for the real-time analysis of time-to-event outcomes. To maximize clinical value, all these approaches must account for specific patient characteristics in order to be able to translate findings into actionable information that can truly guide personalized discussions between patients and their clinicians.

## Implications for clinical practice

Clinicians should be helped to incorporate real-world evidence to guide treatment discussions, as outcomes from clinical practice often differ from those reported in RCTs. For example, patients with PD-L1–negative NSCLC treated with pembrolizumab plus chemotherapy in clinical practice had considerably shorter survival than those treated in the RCT. These findings highlight that treatment benefits seen in RCTs may be less pronounced in certain subgroups and underscore the value of RWE for setting realistic expectations. Besides using RWE, clinicians can also play an important role to motivate patients to record PROMs so that this information can be incorporated in research eventually leading to possibilities to even more make treatment decisions based on outcomes that matters most to patients. Lastly, clinicians and hospitals can help accelerate the generation of RWE by participating in national initiatives such as the DuTOC, which supports standardized, prospective data collection and facilitates prompt insights on treatment outcomes in clinical practice.

**Table 1.** Future directions towards meaningful, and valid outcomes and prompt evidence generation

| | **Lessons learned** | **Future directions** |
|---|---|---|
| **Meaningful outcomes** | OS and PFS remain the most used endpoints in oncology research, but they only partially capture what matters most to patients and face methodological challenges when used to reflect treatment effectiveness in clinical practice. | Future research should incorporate PROM data and promote their integration into EHRs, and clinicians can stimulate patients to generate PROM data. |
| **Valid outcomes** | Observational studies offer complementary evidence to RCTs and support clinical decision making, yet their findings must be interpreted with caution due to potential biases. | Future observational studies should incorporate the principles of target trial emulation to strengthen their validity. |
| **Prompt evidence generation** | Many studies are conducted within a limited number of centers, leading to slow cohort recruitment, and the reliance on manual data collection further delays the prompt generation of evidence. | Future efforts should focus on data harmonization and standardization and use advanced analytic tools (such as text-mining and Bayesian survival modelling) to provide prompt evidence. |

## Conclusion

In conclusion, the findings of this thesis show that outcomes of NSCLC patients treated with immunotherapy in clinical practice often differ from those reported in clinical trials. It further highlights that PFS is difficult to measure consistently and is prone to bias when comparing cohorts, and that methodological advancements such as text mining and Bayesian survival analysis offer opportunities to accelerate real-world outcome generation beyond clinical trials. Ultimately, the value of real-world evidence lies in guiding patients and clinicians in making the best possible treatment decisions.

## References

1. Cramer-van der Welle CM, Verschueren MV, Schramel FMNH, Klungel OH, Groen HJM, Peters BJM, et al. Real-world outcomes versus clinical trial results of immunotherapy in stage IV non-small cell lung cancer in the Netherlands. Sci Rep. 2021;11:1388.
2. Brinkhuis F, Goettsch WG, Mantel-Teeuwisse AK, Bloem LT. High cost oncology drugs without proof of added benefit are burdening health systems. *BMJ*. 2024;384:q511
3. Food and Drug Administration. *Approaches to Assessment of Overall Survival in Oncology Clinical Trials: Draft Guidance for Industry.* Silver Spring (MD): FDA; 2025 Aug 18. Available from: FDA Guidance Documents repository. Docket No. FDA-2024-D-5850
4. Seghers PALN, Wiersma A, Festen S, Stegmann ME, Soubeyran P, Rostoft S, et al. Patient Preferences for Treatment Outcomes in Oncology with a Focus on the Older Patient—A Systematic Review. Cancers (Basel). 2022;14(5):1147.
5. Sherry AD, Msaouel P, Lin TA, Abi Jaoude J, Kouzy R, Beck EJ, et al. Postprogression therapy and confounding for the estimated treatment effect on overall survival in phase III oncology trials. BMJ Oncol. 2024;3:e000322
6. Ković B, Jin H, et al. Evaluating progression-free survival as a surrogate outcome for health-related quality of life in oncology: a systematic review and quantitative analysis. JAMA Intern Med. 2018;178(12):1568-1575.
7. Prasad V, Kim C, Burotto M, Vandross A. The strength of association between surrogate end points and survival in oncology: a systematic review of trial-level meta-analyses. JAMA Intern Med. 2015;175(8):1389-1402.
8. Mittal A, Gyawali B. Frequently asked questions on surrogate endpoints in oncology: opportunities, pitfalls and the way forward. eClinMed. 2024;76:102824.
9. Sherry AD, Miller AM, Parlapalli JP, et al. Overall Survival and Quality-of-Life Superiority in Modern Phase 3 Oncology Trials: A Meta-Epidemiological Analysis. JAMA Oncol. 2025;11(7):718-724.
10. Shrestha S, Martin C, Burton M, et al. Quality of life versus length of life considerations in cancer patients: a systematic review. Psychooncology. 2019;28(7):1367-1380.
11. Weldring T, Smith SM. Patient-reported outcomes (PROs) and patient-reported outcome measures (PROMs). Health Serv Insights. 2013;4:61-68
12. European Medicines Agency. Appendix 2 to the guideline on the evaluation of anticancer medicinal products in man. The use of patient reported outcome (PRO) measures in oncology studies. Scientific guideline. London: EMA; 2016. Doc. Ref. EMA/CHMP/292464/2014.
13. Kyte D, Reeve BB, Efficace F, et al; ISOQOL EMA Response Taskforce. International Society for Quality of Life Research commentary on the draft European Medicines Agency reflection paper on the use of patient-reported outcome (PRO) measures in oncology studies. Qual Life Res. 2016 Feb;25(2):359-362.
14. Teixeira MM, Borges FC, Ferreira PS, Rocha J, Sepodes B, Torre C. A review of patient-reported outcomes used for regulatory approval of oncology medicinal products in the European Union between 2017 and 2020. Front Med (Lausanne).2022;9:968272.
15. Pe M, Alanya A, Falk RS, Amdal CD, Bjordal K, Chang J, et al.; SISAQOL-IMI Consortium. Setting International Standards in Analysing Patient-Reported Outcomes and Quality of Life Endpoints in Cancer Clinical Trials-Innovative Medicines Initiative (SISAQOL-IMI): stakeholder views, objectives, and procedures. Lancet Oncol. 2023 Jun;24(6):e270-e283.
16. de Jong AJ, van Rijssel TI, Zuidgeest MGP, van Thiel GJMW, Askin S, Fons-Martínez J, et al; Trials@Home Consortium. Opportunities and challenges for decentralized clinical trials: European regulators' perspective. Clin Pharmacol Ther. 2022;112(2):344-52

17. de Jong WK, van der Velde FTA, Keijser A, Groen HJM, van Lindert ASR, Vink GR, Hendriks LEL, van den Heuvel MM, et al. **DuTOC:** The prospective Dutch Thoracic Oncology Cohort: A nationwide infrastructure to collect clinical data, PROMs and bio-material to support thoracic oncology studies. Lung cancer. 2025; 208:108728.
18. Derksen JWG, Vink GR, Elferink MAG, Roodhart JML, Verkooijen HM, van Grevenstein WMU, et al; PLCRC Study Group. The Prospective Dutch Colorectal Cancer (PLCRC) cohort: real-world data facilitating research and clinical care. Sci Rep. 2021;11(1):3923.
19. European Health Data Space (EHDS) [internet]. European Commission – Public Health; [cited 2025 Nov 16]. Available from: https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space-regulation-ehds_en
20. Health-RI [Internet]. Utrecht. Stichting Health-RI; [cited 2025 16 nov]. Available from: https://www.health-ri.nl
21. Collins R, Bowman L, Landray M, Peto R. The magic of randomization versus the myth of real-world evidence. N Engl J Med. 2020 Feb 13;382(7):674–678.
22. Al-Sahab B, Leviton A, Loddenkemper T, Paneth N, Zhang B. Biases in electronic health records data for generating real-world evidence: an overview. J Health Inform Res. 2023;8(1):121–139.
23. Background – Progression-Free Survival: What does it mean for psychological well-being or quality of life? Agency for healthcare research and quality (US); 2013;1: 2-8
24. Ackerman B, Gan RW, Meyer CS, Wang JR, Zhang Y, Hayden J, et al. Measurement error and bias in real-world oncology endpoints: challenges in the use of progression-free survival from external data sources. Front Drug Saf Regul. 2024;4;1423493
25. Panageas KS, Ben-Porat L, Dickler MN, Chapman PB, Schrag D. When you look matters: the effect of assessment schedule on progression-free survival. J Natl Cancer Inst. 2007;99(6):428–432. 1423493
26. Dugué AE, Pulido M, Chabaud S, Belin L, Gal J. How to deal with interval-censored data practically while assessing progression-free survival: a step-by-step guide using SAS and R software. Clin Cancer Res. 2016;22(23):5629–5635.
27. Blumenthal GM, Gong Y, Kehl K, Mishra-Kalyani P, Goldberg KB, Khozin S, 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(5):830-838.
28. Selection bias. In: Epidemiology Uninitiated. Chapter 4. Measurement error and bias. BMJ; cited 2025 Aug 31.
29. Nguyen VT, Engleton M, Davison M, Ravaud P, Porcher R, Boutron I. Risk of bias in observational studies using routinely collected data of comparative effectiveness research: a meta-research study. BMC Med. 2021;19(1):297
30. Suissa S. Immortal time bias in pharmacoepidemiology. Am J Epidemiol. 2008 Feb 15;167(4):492-9.
31. Popovic A, Heucker MR. Study Bias [internet]. 2023. [Accessed on: 2025 Aug 31]. Available via: https://www.ncbi.nlm.nih.gov/books/NBK574513/
32. European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP). Guide on methodological standards in pharmacoepidemiology. Chapter 6: Methods to address bias and confounding [Internet]; 2023. [Accessed on: 2025 Aug 31]. Available from: https://encepp.europa.eu/document/download/f6e403a6-8033-4c22-a5ff-195ba3666299_en?filename=01.ENCePPMethodsGuideRev.11.pdf%3E.
33. Freemantle N, Marston L, Walters K, Wood J, Reynolds MR, Petersen I. Making inferences on treatment effects from real world data: propensity scores, confounding by indication, and other perils for the unwary in observational research. BMJ. 2013. 346:f6409
34. Haneuse S. Distinguishing selection bias and confounding by indication in comparative effectiveness research. Med Care. 2016 Jul;54(7):e23-e29.

35. Matthews A, Young JC, Kurth T. The target trial framework in clinical epidemiology: principles and applications. J Clin Epidemiol. 2023 Dec; 164:112-115.
36. Wang SV, Sreedhara SK, Schneeweiss S. Reproducibility of real-world evidence studies using clinical practice data to inform regulatory and coverage decisions. Nat Commun. 2022;13(1):5126.
37. Franklin J, Patorno E, Desai R, Glynn RJ, Martin D, Quinto K. Emulating randomized clinical trials with nonrandomized real-world evidence studies: First results from the RCT DUPLICATE initiative.Circulation. 2021;143(10):1002-1013.
38. Polito L, Liang Q, Pal N, Mpofu P, Sawas A, Humblet O, Rufibach K, Heinzmann D. Applying the estimand and target trial frameworks to external control analyses using observational data: a case study in the solid tumor setting. Front Pharmacol. 2024 Jan 26;15:1223858.
39. Hernán MA, Wamg W, Leaf D. Target trial emulation: A framework for causal inference from observational data. JAMA. 2022;328(24): 2446-2447
40. Porter ME, Lee TH. From volume to value in health care: the work begins. JAMA. 2016;316(10): 1047-8
41. International Consortium for Health Outcomes Measurement (ICHOM). ICHOM Standard Sets. Boston (MA): ICHOM; 2025. Cited 2025 Aug 31. Available from: https://ichom.org/standard-sets/
42. Integraal Kankercentrum Nederland (IKNL). About IKNL. Utrecht: IKNL; 2025. Cited 2025 Aug 31. Available from: https://iknl.nl/
43. The Observational Medical Outcomes Partnership (OMOP) Common Data Model. In: OHDSI .2025. Cited 2025 Aug 31. Available from: https://www.ohdsi.org/data-standardization/
44. Wang L, Wen A, Sunyang F, Xiaoyang R, Huang M, Rui L, et al. A scoping review of OMOP Common Data Model adoption for cancer research using real-world data. NPJ Digit Med. 2025;8(1):189
45. Cow R, Drkulec H, James HBLm, Tsai J, Nafees A, Kumar S, et al. The use of wearable devices in oncology patients: a systematic review. Oncologist. 2023 Nov;29(4):e419-430.

# CHAPTER 5

# Author’s contribution

### Chapter 1 General introduction

I designed the outline of the general introduction, carried out the literature review, and wrote the chapter. Throughout the process, I sought input from my supervisory team and incorporated their feedback to refine the text.

### Chapter 2.1 Real-world effectiveness of durvalumab after concurrent or sequential chemoradiotherapy in patients with stage III NSCLC compared with clinical trial results

I contributed to formulating the research question and designing the study, was responsible for data acquisition, data analysis and visualization, contributed to data interpretation, and wrote and revised the manuscript. During this project, I supervised a MSc student, guiding her in data collection and data analysis, and helping her with writing her MSc thesis. These contributions were performed with input and feedback from the co-authors.

### Chapter 2.2 Pembrolizumab plus chemotherapy per PD-L1 stratum in patients with metastatic non-small cell lung cancer: real-world effectiveness versus trial efficacy.

I contributed to formulating the research question and designing the study, was responsible for data acquisition, data analysis and visualization, contributed to data interpretation, and wrote and revised the manuscript. During this project, I supervised a MSc student, guiding her in data collection and data analysis, and helping her with writing her MSc thesis. These contributions were performed with input and feedback from the co-authors.

### Chapter 2.3 Comparative effectiveness of chemotherapy with nivolumab plus ipilimumab versus chemotherapy with pembrolizumab in PD-L1 negative advanced stage NSCLC patients.

I contributed to formulating the research question and designing the study, was responsible for data acquisition, contributed to data interpretation, and wrote and revised the manuscript. These contributions were performed with input and feedback from the co-authors.

### Chapter 2.4 The association between gut microbiome affecting concomitant medication and the effectiveness of immunotherapy in patients with stage IV NSCLC

I contributed to formulating the research question and designing the study, was responsible for data acquisition, data analysis and visualization, contributed to data interpretation, and wrote and revised the manuscript. These contributions were performed with input and feedback from the co-authors.

### Chapter 3.1 Definition, measurement and reporting of progression-free survival in randomized clinical trials and observational studies of patients with advanced non-small cell lung cancer treated with immunotherapy: a scoping review.

I contributed to formulating the research question and designing the study, and was responsible for the literature search, title/abstract screening, and full-text screening. I performed the data extraction, data analysis, and data visualization, contributed to data interpretation, and wrote and revised the manuscript. During this project, I supervised two MSc students, guiding them in conducting the literature search, extracting data, and writing their MSc theses. These contributions were performed with input and feedback from the co-authors.

### Chapter 3.2 Development and portability of a text-mining algorithm for capturing disease progression in electronic health records of patients with stage IV non-small cell lung cancer treated with immunochemotherapy.

I contributed to formulating the research question and designing the study, was responsible for data acquisition, data analysis and visualization, contributed to data interpretation, and wrote and revised the manuscript. These contributions were performed with input and feedback from the co-authors.

### Chapter 3.3 A Bayesian approach to compare accumulating time-to-event data from clinical practice with RCT data: a case study of overall survival in non-small cell lung cancer patients treated with immunotherapy.

I contributed to formulating the research question and designing the study, was responsible for data acquisition and data visualization. I contributed to discussions on the statistical approach, model development and data interpretation. I wrote and revised the manuscript. These contributions were performed with input and feedback from the co-authors.

### Chapter 4 General discussion

I designed the outline of the general discussion, carried out the literature review, and wrote the chapter. Throughout the process, I sought input from my supervisory team and incorporated their feedback to refine the text. The general discussion represents my opinion and view.

# CHAPTER 6

## Summaries

## English summary

The introduction of immune checkpoint inhibitors (ICIs) has altered the treatment landscape of advanced non-small cell lung cancer (NSCLC) without targetable driver mutations, shifting the standard of care from chemotherapy to immunotherapy-based strategies. Evidence supporting the use of ICIs is primarily derived from randomized controlled trials (RCTs), which evaluate treatment effects under controlled conditions in selected patient populations. However, outcomes observed in RCTs may differ from those observed in clinical practice due to differences in patient characteristics, treatment adherence, and follow-up assessments. Real-world evidence can complement RCT evidence by providing insight into outcomes in more heterogeneous populations and settings. As new immunotherapy-based treatments are rapidly introduced, there is an increasing need for timely evaluation of their effectiveness in clinical practice. Nevertheless, generating prompt and reliable real-world evidence remains challenging. Therefore, this thesis aimed to complement data on survival outcomes reported in RCTs of patients with advanced NSCLC with data from patients treated with immunotherapy in clinical practice, while also evaluating methodological challenges and advancements for prompt evidence generation.

### Outcomes of immunotherapy in lung cancer patients treated in clinical practice

In **Chapter 2.1**, the effectiveness of durvalumab following concurrent or sequential chemoradiotherapy in patients with stage III NSCLC was investigated. The findings showed that durvalumab improved both progression-free survival (PFS) and overall survival (OS) in clinical practice as compared with historical controls treated with chemoradiotherapy alone. The median PFS observed in clinical practice was substantially longer than the median PFS reported in the PACIFIC trial, whereas OS did not differ. This difference in PFS may be explained by differences in follow-up and outcome assessment between clinical practice and the clinical trial. In **Chapter 2.2**, the effectiveness of first-line pembrolizumab combined with chemotherapy in patients with metastatic NSCLC was examined, stratified by PD-L1 expression. Patients with non-squamous metastatic NSCLC and PD-L1 expression <1% who were treated with pembrolizumab plus chemotherapy in clinical practice had statistically significantly shorter OS than those in the corresponding RCT. No statistically significant differences were observed in the other PD-L1 strata. In **Chapter 2.3**, the outcomes of patients with metastatic NSCLC and PD-L1 expression <1% treated with nivolumab plus ipilimumab combined with chemotherapy in clinical practice were compared with those of historical controls treated with pembrolizumab plus chemotherapy. Patients treated with nivolumab plus ipilimumab showed numerically better treatment responses and slightly longer PFS than those treated with pembrolizumab plus chemotherapy. These findings suggest that nivolumab plus ipilimumab may be more effective than pembrolizumab plus chemotherapy in clinical practice for this patient population, although longer follow-up is needed to assess the impact on overall survival. In **Chapter 2.4**, the potential effect of concomitant medication on the effectiveness of immunotherapy in patients with advanced NSCLC was investigated, as certain medications such as antibiotics have been hypothesized to influence immunotherapy outcomes through alterations of the gut microbiome. To address potential confounding by indication, outcomes in patients treated with immunotherapy were compared with those of a matched historical cohort treated with chemotherapy. Patients who used antibiotics or opioids showed worse survival outcomes in both cohorts. This finding suggests that the observed association is more likely related to underlying patient characteristics rather than to a direct effect of these medications on immunotherapy effectiveness.

### Methodological challenges and advancements for prompt evidence generation

In **Chapter 3**, methodological challenges and advancements related to prompt evidence generation were addressed. In **Chapter 3.1**, a scoping review was conducted to investigate how PFS is defined, measured, and reported for RCTs and observational studies of patients with advanced NSCLC treated with immunotherapy. The review showed that, compared with RCTs, observational studies frequently lack clear reporting on PFS definitions and measurement characteristics and that, when reported, these vary widely across studies. This variability hinders comparisons of PFS between observational studies and between observational studies and clinical trials, highlighting the need for careful interpretation of real-world PFS outcomes. In **Chapter 3.2**, the development and evaluation of a text-mining algorithm to capture disease progression from electronic health record data of patients with metastatic NSCLC treated with immunochemotherapy were addressed. The algorithm was developed in two Dutch hospitals and subsequently applied in other Dutch hospitals to assess its portability across different clinical settings. The results showed that disease progression could be detected accurately over time across hospitals. PFS estimates derived from text mining were similar to those obtained through manual chart review, indicating that this approach is suitable for prospective and scalable follow-up of disease progression in multicenter real-world cohorts. In **Chapter 3.3**, a Bayesian survival model was developed to enable early comparison of time-to-event outcomes between accumulating real-world data from clinical practice and fixed RCT data. The aim was to provide rapid and easily interpretable insights into potential differences in treatment effectiveness to support timely decision-making. When applied to two lung cancer datasets, the model showed that, provided an adequate model fit was achieved, reliable estimates and clear evidence of differences in survival outcomes could be obtained before data collection for typical observational studies was complete.

## Discussion

The discussion in **Chapter 4** places the findings presented in this thesis in a broader perspective and addresses three key themes: the need for meaningful outcomes, valid outcomes, and prompt evidence generation. First, the theme of meaningful outcomes. While OS and PFS remain central outcomes for decision-making about oncology medicines, they do not always capture what matters most to patients, including their quality of life. Future research should therefore include outcomes that are truly meaningful for patients with advanced NSCLC, which requires the systematic collection and use of patient-reported outcome measures. Second, the theme of valid outcomes. Real-world data are susceptible to biases inherent to observational research, including information bias, selection bias, and confounding by indication. The studies in this thesis illustrate how explicitly accounting for these sources of bias through harmonized outcome definitions and appropriate comparison strategies can improve the validity of comparisons with RCT and other studies. Looking ahead, future methodological research should further strengthen the validity of observational studies of new cancer medication, for example by exploring approaches such as target trial emulation. Third, the theme of prompt evidence generation. As treatment options evolve rapidly, timely insight into real-world effectiveness is essential but is often constrained by slow data collection and limitations of conventional analytical approaches. Addressing these challenges requires scalable data infrastructures, improved data harmonization across hospitals, and methods that support continuous or near–real-time analysis. Methodological advancements such as text mining of unstructured clinical data and Bayesian survival modeling offer promising opportunities to accelerate and strengthen real-world evidence generation.

In conclusion, this thesis showed that survival outcomes of patients with advanced NSCLC treated with immunotherapy in clinical practice differ from those reported in RCTs, highlighting the added value of real-world data in complementing RCT evidence. It further highlighted that PFS is difficult to measure consistently in real-world settings and is prone to bias, whereas methodological advancements, such as text mining and Bayesian survival modeling, can enable more prompt generation of real-world evidence.

## Dutch summary / Nederlandse samenvatting

De komst van immuuntherapie heeft de behandeling van uitgezaaide niet-kleincellige longkanker (NSCLC) ingrijpend veranderd. In het verleden kregen longkankerpatiënten vooral chemotherapie. Tegenwoordig is immuuntherapie een hoeksteen in de standaardbehandeling. De kennis over deze behandelingen komt vooral voort uit gerandomiseerde klinische studies. Dit zijn onderzoeken waarin een nieuwe behandeling wordt vergeleken met een bestaande behandeling of placebo. Dit type onderzoek gebeurt onder streng gecontroleerde omstandigheden en bij zorgvuldig geselecteerde patiënten, bijvoorbeeld patiënten die relatief fit zijn en verder geen andere aandoeningen hebben. Zo kan op een zo betrouwbaar mogelijke manier worden vastgesteld wat het effect van de nieuwe behandeling is ten opzichte van de bestaande behandeling. In de dagelijkse praktijk (‘real world’) kan de behandeling van longkankerpatiënten echter andere resultaten geven dan de resultaten van klinische studies. Dit komt mede omdat patiënten in de dagelijkse praktijk geen perfecte afspiegeling vormen van de zorgvuldig geselecteerde patiëntengroep die onderdeel uitmaakt van een klinische studie. Zo zijn patiënten buiten studieverband vaak ouder, hebben zij naast longkanker mogelijkerwijs ook andere aandoeningen en krijgen zij tijdens hun behandeling vaak minder controles. In klinische studies worden patiënten namelijk intensief gecontroleerd, bijvoorbeeld met extra afspraken, scans en bloedonderzoeken. Onderzoek met gegevens uit de dagelijkse praktijk kan daarom een belangrijke aanvulling zijn op klinische studies om de effectiviteit en veiligheid van immuuntherapiebehandelingen te bepalen. Het bewijs dat uit dit type onderzoek voortkomt, wordt ook wel real-world evidence genoemd.

Omdat er steeds nieuwe immuuntherapieën beschikbaar komen en behandelingen ook samen worden gebruikt, is het belangrijk om snel duidelijkheid te krijgen over hun werkzaamheid en veiligheid in de praktijk. Zonder tijdig inzicht kan een niet effectieve behandeling in de praktijk langer worden gebruikt dan wenselijk is, terwijl er mogelijk betere, minder toxische alternatieven beschikbaar zijn. Het verzamelen van betrouwbare gegevens over behandeluitkomsten bij patiënten die in de dagelijkse praktijk worden behandeld is echter complex en tijdrovend. Veel informatie staat namelijk in uitgeschreven notities van artsen, vaak met medisch jargon en niet altijd eenduidig beschreven, waardoor het moeilijk is om deze gegevens snel te analyseren. Het doel van dit proefschrift was om de resultaten uit klinische studies bij patiënten met NSCLC die met immuuntherapie werden behandeld aan te vullen met gegevens uit de dagelijkse praktijk van patiënten met dezelfde ziekte en behandeling. Daarnaast is onderzocht hoe dit soort praktijkbewijs sneller en zorgvuldiger kan worden verzameld en geanalyseerd.

## Behandeluitkomsten van immuuntherapie in de dagelijkse praktijk

In **Hoofdstuk 2** zijn de behandeluitkomsten van immuuntherapie onderzocht bij patiënten met gevorderde NSCLC in de dagelijkse praktijk. In **Hoofdstuk 2.1** werd onderzocht hoe effectief durvalumab is na een behandeling met chemotherapie en radiotherapie bij patiënten met lokaal gevorderde NSCLC (stadium III). De uitkomsten werden vergeleken met de uitkomsten van patiënten die alleen chemotherapie en radiotherapie kregen en met de uitkomsten van patiënten die dezelfde behandeling kregen binnen een klinische studie. Patiënten die in de dagelijkse praktijk met durvalumab werden behandeld, leefden langer zonder dat de ziekte verergerde en leefden over het algemeen langer (‘totale overleving’) dan vergelijkbare patiënten die alleen chemotherapie en radiotherapie kregen. Opvallend was dat het bij patiënten in de dagelijkse praktijk langer duurde voordat de ziekte verergerde dan bij patiënten in de klinische studies, terwijl de totale overleving vergelijkbaar bleef. Een mogelijke verklaring hiervoor is dat patiënten in de dagelijkse praktijk minder vaak en consistent worden gecontroleerd dan in een klinische studie, waardoor verschillen kunnen ontstaan in het moment waarop verergering van de ziekte wordt vastgesteld.

In **hoofdstuk 2.2** werd onderzocht hoe goed pembrolizumab plus chemotherapie werkt bij patiënten met uitgezaaide NSCLC (stadium IV) in de dagelijkse praktijk en werden deze resultaten vergeleken met die uit klinische studies. De uitkomsten werden geanalyseerd bij patiënten met verschillende niveaus van PD-L1, een eiwit op de tumorcellen dat het immuunsysteem kan onderdrukken, waarbij een hogere PD-L1-expressie meestal samenhangt met een beter effect van immuuntherapie. De resultaten lieten zien dat patiënten zonder PD-L1-expressie in de dagelijkse praktijk korter leefden dan patiënten in de klinische studie. Bij patiënten met een hogere PD-L1-expressie werden geen duidelijke verschillen gevonden. In **Hoofdstuk 2.3** werden twee behandelingen bij patiënten zonder PD-L1-expressie met elkaar vergeleken: een combinatie van nivolumab en ipilimumab plus chemotherapie (NIC) en een combinatie van pembrolizumab plus chemotherapie (PC). Patiënten die NIC kregen, hadden iets vaker een duidelijke afname van de grootte van hun tumor dan patiënten die PC kregen. Ook duurde het langer voordat hun ziekte verergerde. Dit suggereert dat NIC voor deze patiënten in de dagelijkse praktijk mogelijk beter werkt dan PC. Of dit ook betekent dat ze langer leven, moet nog verder worden onderzocht. In **Hoofdstuk 2.4** werd onderzocht of geneesmiddelen die gelijktijdig werden gebruikt, zoals antibiotica of opioïden, invloed hebben op het effect van immuuntherapie, mogelijk door de verstoring van darmbacteriën (microbioom) die het immuunsysteem ondersteunen. Hiervoor werden patiënten met NSCLC die immuuntherapie kregen vergeleken met een vergelijkbare groep patiënten die eerder met chemotherapie waren behandeld. In beide groepen leefden patiënten die antibiotica of opioïden gebruikten gemiddeld korter. Dit laat zien dat de slechtere uitkomsten ook beïnvloed zijn door de onderliggende gezondheid van de patiënt zelf en niet direct (of slechts) met een verstoring van het microbioom door deze medicijnen.

## Methodologische uitdagingen en oplossingen voor sneller bewijs

**Hoofdstuk 3** richtte zich op de methodologische uitdagingen en mogelijke oplossingen om snel betrouwbaar bewijs te genereren bij patiënten met gevorderde NSCLC in de dagelijkse praktijk. In **hoofdstuk 3.1** werd onderzocht hoe de tijd tot ziekteverergering, ook wel progressievrije overleving genoemd, wordt gedefinieerd en gemeten voor klinische studies en in observationele studies. Hieruit bleek dat voor observationele studies vaak onvoldoende duidelijk beschreven wordt hoe deze uitkomst is vastgesteld. Ook bestaan er grote verschillen in de definitie en manier van meten van progressievrije overleving. Daardoor is het lastig om resultaten van verschillende observationele studies, en van observationele studies en klinische studies, goed met elkaar te vergelijken. In **hoofdstuk 3.2** werd een tekst-mining algoritme ontwikkeld dat automatisch ziekteprogressie kan herkennen uit de vrije teksten in de medische dossiers van longkankerpatiënten. Het algoritme werd getest in meerdere Nederlandse ziekenhuizen en bleek goed te werken: de uitkomsten kwamen grotendeels overeen met handmatige dossierbeoordeling. De resultaten laten zien dat dit tekst-mining algoritme geschikt is om patiënten in meerdere ziekenhuizen in de tijd te volgen en zo bijna real-time inzicht te krijgen in het moment waarop ziekteprogressie optreedt. In **Hoofdstuk 3.3** werd een statistisch model ontwikkeld dat vaste resultaten uit klinische studies vergelijkt met nieuwe gegevens die stap voor stap binnenkomen uit de dagelijkse praktijk. Met dit model kan al in een vroeg stadium worden beoordeeld of een nieuw geneesmiddel in de praktijk dezelfde resultaten oplevert als in de klinische studie. Zo kan sneller worden gesignaleerd wanneer een behandeling in de praktijk minder effectief blijkt dan verwacht, waardoor de behandelrichtlijnen tijdig aangepast kunnen worden.

## Beschouwing

In **Hoofdstuk 4** werden de resultaten van dit proefschrift in een breder perspectief geplaatst. Hierbij is gekeken naar drie thema's die vaak als belangrijke parameters worden beschouwd bij het beoordelen van nieuwe behandelingen: betekenisvolle uitkomsten, betrouwbare uitkomsten en het snel leveren van bewijs. Deze thema's bieden een kader om de resultaten te interpreteren en te bespreken in de context van de dagelijkse praktijk. Ten eerste het thema betekenisvolle uitkomsten. De meeste onderzoeken bij longkankerpatiënten zich vooral richten op de totale overleving en de progressievrije overleving. Deze uitkomsten zeggen echter niet automatisch iets over hoe iemand zich voelt, inclusief diens kwaliteit van leven. Toekomstig onderzoek zou daarom vaker gebruik moeten maken van vragenlijsten waarin patiënten zelf aangeven hoe zij zich voelen en hoe zij in het dagelijks leven functioneren. Ten tweede, het thema betrouwbare uitkomsten. Onderzoek op basis van gegevens uit de dagelijkse praktijk is gevoelig voor vertekening, bijvoorbeeld door verschillen tussen studies in het type patiënt dat wordt meegenomen of door verschillen in meetmethoden. Door hier vooraf expliciet rekening mee te houden en uitkomstdefinities beter op elkaar af te stemmen, kunnen resultaten uit de praktijk betrouwbaarder worden vergeleken met andere studies, zowel klinische studies als studies met gegevens uit de dagelijkse praktijk. Ten derde, het thema snel bewijs leveren. Snelle informatie over de effectiviteit van nieuwe behandelingen is essentieel om onvoldoende effectieve therapieën in de dagelijkse praktijk tijdig te signaleren. Helaas zijn deze praktijkuitkomsten meestal niet snel beschikbaar, omdat gegevens uit meerdere ziekenhuizen moeten worden verzameld om betrouwbaar te zijn. Daarnaast staat veel belangrijke informatie in patiëntendossiers in vrije tekst staat, waardoor het verkrijgen van uitkomsten langer duurt. Daarom zijn grootschalige, goed op elkaar afgestemde datasystemen nodig, gecombineerd met slimme analysemethoden, zodat praktijkgegevens sneller en beter benut kunnen worden.

## Conclusie

Dit proefschrift laat zien dat behandeluitkomsten van immuuntherapie bij patiënten met gevorderde NSCLC in de dagelijkse praktijk afwijken van de resultaten uit klinische studies, en onderstreept daarmee de meerwaarde van gegevens uit de dagelijkse praktijk als aanvulling op klinische studies. Daarnaast blijkt dat progressievrije overleving in de praktijk lastig en niet altijd betrouwbaar te meten is. Nieuwe methoden, zoals automatische analyse van medische dossiers en geavanceerde statistische modellen, maken het mogelijk om sneller en beter te begrijpen hoe behandelingen in de dagelijkse praktijk werken en kunnen daarmee een waardevolle toevoeging zijn aan de bestaande kennis over de effectiviteit van immuuntherapie.

# CHAPTER 7

## Appendices

## List of author affiliations

At time of publications

Hanieh Abedien Kalkhoran
Department of Clinical Pharmacy and Toxicology, Leiden University Medical Center, Leiden, the Netherlands; Department of Pharmacy, Haga Teaching Hospital, The Hague, the Netherlands.

Annette Bijsmans
Department of Pulmonology, Maasstad Hospital, Rotterdam, Netherlands.

Lourens Bloem
Division of Pharmacoepidemiology & Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, The Netherlands.

Ben van den Borne
Department of Respiratory Medicine, Catharina Hospital, Eindhoven, the Netherlands.

Christine Cramer- van der Welle
Santeon Hospital Group, Utrecht, The Netherlands.

Sander Croes
Department of Clinical Pharmacy & Toxicology, Maastricht University Medical Center+, Maastricht, The Netherland; CARIM School for Cardiovascular Disease, Maastricht University Medical Center+, Maastricht, The Netherlands

Maarten Deenen
Department of Clinical Pharmacy and Toxicology, Leiden University Medical Center, Leiden, the Netherlands; Department of Clinical Pharmacy, Catharina Hospital, Eindhoven, the Netherlands.

Talitha Dijs
Department of Clinical Pharmacy, St. Antonius Hospital, Utrecht/Nieuwegein, The Netherlands

Toine Egberts
Division of Pharmacoepidemiology & Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, The Netherlands; Department of Clinical Pharmacy, University Medical Center Utrecht, Utrecht, The Netherlands.

Ewoudt van de Garde
Division of Pharmacoepidemiology & Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, The Netherlands; Department of Clinical Pharmacy, St. Antonius Hospital, Utrecht/Nieuwegein, The Netherlands.

Judith Gulikers
Department of Clinical Pharmacy & Toxicology, Maastricht University Medical Center+, Maastricht, The Netherlands; CARIM School for Cardiovascular Disease, Maastricht University Medical Center+, Maastricht, The Netherlands

Lisa Hendriks
Department of Respiratory Medicine, Maastricht University Medical Centre, GROW School for Oncology & Developmental Biology, Maastricht, The Netherlands.

Dagmar Hiensch
Department of Clinical Pharmacy, St. Antonius Hospital, Utrecht/Nieuwegein, The Netherlands.

Lisanne Kastelijn
Department of Pulmonary Diseases, St. Antonius Hospital, Utrecht/Nieuwegein, The Netherlands.

Miriam Koopman
Department of Medical Oncology, University Medical Centre Utrecht, Utrecht University, Utrecht, The Netherlands

Veerle Kruik
Department of Clinical Pharmacy, St. Antonius Hospital, Utrecht/Nieuwegein, The Netherlands.

Bas Peters
Department of Clinical Pharmacy, St. Antonius Hospital, Utrecht/Nieuwegein, The Netherlands; Division of Pharmacoepidemiology & Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, The Netherlands.

Peter Plomp
Pulmonology Department, Isala Hospital, Zwolle, The Netherlands.

Franz Schramel
Department of Pulmonary Diseases, St. Antonius Hospital, Utrecht/Nieuwegein, The Netherlands.

Jeroen Schuurkamp
Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, The Netherlands

Arthur Smit
Department of Pulmonary Diseases, OLVG, Amsterdam, The Netherlands.

Vassiliki Tassopoulou
Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, The Netherlands.

Merel Tonn
Department of Clinical Pharmacy, St. Antonius Hospital, Utrecht/Nieuwegein, The Netherlands.

Eline Uitvlugt
Department of Clinical Pharmacy, OLVG, Amsterdam, Netherlands.

Robert Visscher
Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, The Netherlands.

Loes Visser
Department of Pharmacy, Haga Teaching Hospital, The Hague, the Netherlands; Department of Hospital Pharmacy, Erasmus Medical Centre, Rotterdam, the Netherlands; Department of Epidemiology, Erasmus Medical Centre, Rotterdam, the Netherlands.

Ard van Veelen
Division of Pharmacoepidemiology & Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, The Netherlands; CARIM School for Cardiovascular Disease, Maastricht University Medical Center+, Maastricht, The Netherland

Daniel Verschueren
Department of Chemistry, KU Leuven-University of Leuven, Leuven, Belgium.

Julliette Zwaveling
Department of Clinical Pharmacy and Toxicology, Leiden University Medical Center, Leiden, the Netherlands.

# Dankwoord

Dit boek is af! Tijd om iedereen te bedanken voor alle inzichten, steun en betrokkenheid in de afgelopen jaren.

Beste prof. dr. Ewoudt van de Garde, wat bijzonder om je nu als professor te mogen aanspreken, een titel die meer dan verdiend is. Als ik terugkijk op de eerste acht jaar van mijn carrière, heb jij daarin een grote rol gespeeld. Dank voor de AIOS-plek in het St. Antonius Ziekenhuis en voor het mogelijk maken van mijn promotietraject. Ook hielp je mij bij de keuze voor de epidemiologie. Ik heb grote bewondering voor jouw snelheid van denken en werken. Waar ik soms nog mijn gedachten probeer te ordenen, heb jij de kern al geformuleerd en een plan van aanpak klaarliggen. Bij jou heb ik geleerd hoe overzicht eruitziet en ook hoe een lege inbox eruit kan zien. Dank je wel voor je begeleiding, je scherpte en voor alles wat je in deze belangrijke jaren voor mij hebt betekend.

Beste prof. dr. Toine Egberts, voordat mijn promotietraject begon kenden wij elkaar nog niet. Via de wandelgangen had ik al veel gehoord. Een indrukwekkende carrière, hoofd van de GSLS en daarnaast ultramarathons rennen in de bergen. Dat schetste het beeld van iemand met een indrukwekkend tempo. Gelukkig bleek diezelfde energie zich ook te vertalen naar betrokken en toegankelijke begeleiding. Ik heb veel van je geleerd. Je liet mij het belang zien van een sterke methodologische basis en het zorgvuldig formuleren van definities. We hebben soms lang over definities gediscussieerd, zo lang dat ik dacht dat ik het wel wist. Maar juist daar zat de essentie. Ook ben je voor mij een voorbeeld in wetenschappelijke integriteit. Zorgvuldig, kritisch en onafhankelijk, met oog voor wat inhoudelijk klopt.

Beste dr. Bas Peters, wat een feest om zeven jaar met jou samen te werken in het St. Antoniusziekenhuis en later ook als copromotor. In die jaren heb ik veel van je geleerd. Klinisch-inhoudelijk, epidemiologisch, maar ook van jouw talent om processen te verbeteren, samenwerkingen te laten floreren en vooruit te kijken. Dank voor al jouw inzichten en voor de vele waardevolle gesprekken over onderzoek, werk en alles daarbuiten. Ook wil ik je bedanken voor de momenten waarop jij mij uit de put wist te praten. Jouw enthousiasme en optimisme zijn aanstekelijk en maken jou een super fijne supervisor en collega. Ik hoop dat we in de toekomst blijven samenwerken en nog veel onderzoek samen zullen doen.

Beste dr. Lourens Bloem, Ik heb veel bewondering voor jouw epidemiologische kennis, ambitie en drive. Als copromotor gaf je daarmee het goede voorbeeld. Altijd met veel projecten tegelijk bezig, soms een nachtje doorwerken en toch steeds werk van hoge kwaliteit afleveren. Dank voor alles wat je mij hebt geleerd en voor de fijne samenwerking en leuke congressen. Hoe taai de inhoud soms ook was, bijvoorbeeld het mathematisch uitwerken van een Bayesiaans survivalmodel (waar we gelukkig het orakel Daniel konden gebruiken), of hoe scherp de deadline ook stond, er was altijd ruimte voor een grap en een goed gesprek. Vooral jouw steun in de laatste, voor mij moeilijke fase van dit proefschrift waardeer ik enorm.

Graag wil ik alle coauteurs bedanken voor alle waardevolle discussies, suggesties en waardevolle bijdragen aan dit proefschrift.

Graag wil ik al mijn fijne collega's van de eenheid farmacie uit het St Antoniusziekenhuis bedanken. Ik ben heel dankbaar voor alles wat jullie mij de afgelopen 7 jaar geleerd hebben. Ik heb mij altijd heel erg op mijn plek gevoeld bij jullie in de vakgroep. Met veel van jullie heb ik nog steeds contact en ik hoop dat dit ook in de toekomst nog zal blijven.

Ik wil ook graag al mijn fijne collega's uit het Erasmus MC ziekenhuis bedanken. Ik ben heel dankbaar voor jullie steun en betrokkenheid bij de afronding van mijn promotietraject. En specifiek wil ik nog Rose en Hugo bedanken. Ik waardeer het enorm dat jullie mij de ruimte hebben gegeven om aan mijn proefschrift te werken toen het privé niet mee zat.

Ik wil ook al mijn dierbare vrienden (zij die het aangaat weten wie ik bedoel) en schoonfamilie bedanken voor alle keren dat jullie geduldig naar mijn promotieverhalen met vakjargon wilden luisteren. Tijdens de afrondende fase van dit proefschrift en na het overlijden van mijn moeder heb ik veel steun aan jullie gehad. Er was altijd ruimte om mijn verdriet te delen en er was altijd wel iemand die op Clemens wilde passen zodat ik aan mijn proefschrift kon werken. Ik ben enorm dankbaar voor jullie steun.

Lieve Haas, Marjo en Lies, jullie wil ik in het bijzonder bedanken. Jullie waren er altijd en hebben het verdriet een stukje dragelijker gemaakt.

Lieve familie, ik weet niet goed hoe ik moet beginnen. Het maakt mij verdrietig dat onze lieve moeder Willy al het moois van het leven niet meer mag meemaken. Ze had nog een grote wens om bij de verdediging van mijn proefschrift te zijn. Helaas heeft de ziekte haar ingehaald. Ik weet zeker dat ze er in gedachten bij is (en vast ook de nodige kritiek zal leveren;)). Jos, Suzanne, Daniel, Henk en Iris, ik ben heel blij met jullie. Het afgelopen jaar was heftig en verdrietig, maar het heeft mij ook extra laten beseffen hoe dankbaar ik ben dat we elkaar hebben. Jullie hebben mij door de laatste fase van dit proefschrift heen gesleept. En Daniel en Suzanne, bedankt dat jullie mijn paranimfen willen zijn. Ik hoop dat Suzanne de traditie van familieparanimfen voortzet. Dan zijn we straks alle drie gepromoveerd en hebben we elkaar als paranimf bijgestaan. Ik weet zeker dat Willy enorm trots zou zijn!

En tot slot, de liefde van mijn leven. Lieve Maarten, je bent er altijd voor mij en was een enorme steun bij het afronden van dit proefschrift. Jouw liefde, relativeringsvermogen en humor maken mij elke dag gelukkig. En met onze kleine grappenmaker Clemens is het leven nog leuker.

# List of publications

## Publications in this thesis

MV Verschueren, T Dijs, JL Gulikers, A van Veelen, S Croes, LEL Hendriks, AAJ Smit, LT Bloem, ACG Egberts, EMW van de Garde and BMJ Peters. Real-world effectiveness of durvalumab after concurrent or sequential chemoradiotherapy in patients with stage III NSCLC compared with clinical trial results. Immunotherapy. 2023 Aug;15(11):839-851.

MV Verschueren, BJM Peters, LT Bloem, VR Kruik, EB Uitvlugt, AR Bijsmans, ACG Egberts and EMW van de Garde. Pembrolizumab plus chemotherapy per PD-L1 stratum in patients with metastatic non-small cell lung cancer: real-world effectiveness versus trial efficacy. Clin Lung Cancer. 2024 Mar;25(2):119-127.

MV Verschueren*, DTA Hiensch*, PMJ Plomp, LA Kastelijn, EMW van de Garde and BJM Peters. Comparative effectiveness of chemotherapy with nivolumab plus ipilimumab versus chemotherapy with pembrolizumab in PD-L1 negative advanced stage NSCLC patients. Clin Lung Cancer. 2025 Jan; 26(4):243-248

MV Verschueren, CM Cramer – van der Welle, M Tonn, FHNH Schramel, BJM Peters and EMW van de Garde. The association between gut microbiome affecting concomitant medication and the effectiveness of immunotherapy in patients with stage IV NSCLC. Sci Rep. 2021 Dec 2;11(1):23331.

MV Verschueren, VP Tassopoulou, R Visscher, J Schuurkamp, BJM Peters, M Koopman, EMW van de Garde, ACG Egberts and LT Bloem. Definition, measurement and reporting of progression-free survival in randomized clinical trials and observational studies of patients with advanced non-small cell lung cancer treated with immunotherapy: a scoping review. ESMO real-world data and digital oncology. 2025 March:7:100118.

MV Verschueren, H Abedian Kalkhoran, M Deenen, BEEM van den Borne, J Zwaveling, LE Visser, LT Bloem, BJM Peters and EMW van de Garde. Development and portability of a text-mining algorithm for capturing disease progression in electronic health records of patients with stage IV non-small cell lung cancer treated with immunochemotherapy. JCO Clin Cancer Inform. 2024 Oct 8:e2400053.

MV Verschueren*, DV Verschueren*, EMW van de Garde and LT Bloem. A Bayesian approach to compare accumulating time-to-event data from clinical practice with RCT data: a case study of overall survival in non-small cell lung cancer patients treated with immunotherapy. CPT Pharmacometrics Syst Pharmacol. 2025 Dec; 14(12): 2006-2013.

** Authors contributed equally*

## Other publications

MV Verschueren, FGA Jansman, DJ Touw and C Kramers. Treating poisoning; how do you choose the best type of gastrointestinal decontamination? Ned Tijdschr Geneesk. 2018. 2018:162:D1574.

CM Cramer-van der Welle, MV Verschueren, M Tonn, BJM Peters, FMNH Schramel, OH Klungel et al. Real-world outcomes versus clinical trial results of immunotherapy in stage IV non-small cell lung cancer (NSCLC) in the Netherlands. Sci Rep. 2021 Mar; 11(1):6306

MPH van den Broek, MV Verschueren and CAJ Knibbe. Critical appraisal of evidence for anti-Xa monitoring and dosing of low-molecular-weight heparin in renal insufficiency. Expert Rev of Clin Pharmacol. 2022 Oct;15(10):1155-1163.

A van Veelen, GDM Veerman, MV Verschueren, JL Gulikers, CMJ Steendam, AJWM Brouns et al. Exploring the impact of patient-specific clinical features on osimertinib effectiveness in a real-world cohort of patients with EGFR mutated non-small cell lung cancer. Int J Cancer. 2024 jan; 154(2):332-342.

H Abedian Kalkhoran, MV Verschueren, BJM Peters, EMW van de Garde, Egbert F Smit, Stefan Bohringer et al. Real-World Effectiveness of Durvalumab in Stage III Unresectable Non-small Cell Lung Cancer: What is the Role of PD-L1 Expression? Clin Lung Cancer. 2025 Nov;26(7):e599-e608.e2.

EJ van Dijk, MM. Smeenk, F Bensch, AM Sadowksa, MV Verschueren, M. Verhaer et al. Checkpoint inhibitor effectiveness after corticosteroids and second-line immunosuppressants for immune-related adverse events in non-small-cell lung cancer. ESMO Open. 2026 Jan.

JCM Prick, MV Verschueren, IA Deijle, SM van Schaik, R Dahmen, PJAM. Brouwers et al. Shared Decision-Making Supported by Outcome Information During Discharge Planning of Patients Hospitalized With Stroke. Neurology Clinical Practice. 2026 Feb.

## Curriculum vitae

Marjon Verschueren was born on February 13, 1991, in Nijmegen, the Netherlands. After graduating from the Nijmeegse scholengemeenschap Groenewoud in 2009, she started her studies in Pharmacy at the University of Groningen. During her Master's program, she completed a research internship at the INSERM Institute in Paris, where she worked on identifying targets in liver fibrosis using a mouse model. After obtaining her PharmD degree in 2016, she worked as a hospital pharmacist at Erasmus MC for one year. She subsequently started her residency in hospital pharmacy at St. Antonius Hospital in Nieuwegein. During her residency, she also obtained a post-master's degree in Epidemiology. Following the completion of her hospital pharmacy training, she started her PhD research at the Division of Pharmacoepidemiology and Clinical Pharmacology at Utrecht University and the St. Antonius Hospital. Her PhD was conducted under the supervision of Prof. Dr. Ewoudt van de Garde, Prof. Dr. Toine Egberts, Dr. Bas Peters, and Dr. Lourens Bloem. As of December 2024, she has been working as a hospital pharmacist at Erasmus MC.