Computational and Structural Biotechnology Journal 24 (2024) 611–621

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

Research article

# Assessment of NSCLC disease burden: A survival model-based meta-analysis study

Nataliya Kudryashova [a,c], Boris Shulgin [a], Nikolai Katuninks [a], Victoria Kulesh [a,b], Gabriel Helmlinger [d], Kirill Zhudenkov [a,b,*], Kirill Peskov [a,b,e]

[a] I.M. Sechenov First Moscow State Medical University, Moscow, Russia
[b] Modeling & Simulation Decisions FZ-LLC, Dubai, UAE
[c] Semenov Research Center of Chemical Physics, Moscow 119991, Russia
[d] Quantitative Medicines LLC, Lexington, MA 02420, USA
[e] Russia Sirius University of Science and Technology, Sirius, Russia

ARTICLE INFO

*Keywords:*
NSCLC disease burden
Model-based meta-analysis
Survival model
Oncology
QALY
Life years gained

ABSTRACT

We present a meta-analytics approach to quantify NSCLC disease burden by integrative survival models. Aggregated survival data from public sources were used to parameterize the models for early as well as advanced NSCLC stages incorporating chemotherapies, targeted therapies, and immunotherapies. Overall survival (OS) was predicted in a heterogeneous patient cohort based on various stratifications and initial conditions. Pharmacoeconomic metrics (life years gained (LYG) and quality-adjusted life years (QALY) gained), were evaluated to quantify the benefits of specialized treatments and improved early detection of NSCLC. Simulations showed that the introduction of novel therapies for the advanced NSCLC sub-group increased median survival by 8.1 months (95 % CI: 5.9, 10.0), with corresponding gains of 2.9 months (95 % CI: 2.2, 3.6) in LYG and 1.65 months (95 % CI: 1.2, 2.0) in QALY. Scenarios representing improved detection of early cancer in the whole patient cohort, revealed up to 17.6 (95 % CI: 16.5, 19.0) and 15.7 months (95 % CI: 14.8, 16.6) increase in median survival, with respective gains of 6.2 months (95 % CI: 5.9, 6.4) and 5.2 months (95 % CI: 4.9, 5.4) in LYG and 6.6 months (95 % CI: 6.4, 6.7) and 6.0 months (95 % CI: 5.9, 6.2) in QALY for conventional and optimal treatment. This integrative modeling platform, aimed at characterizing cancer burden, allows to precisely quantify the cumulative benefits of introducing specialized therapies into the treatment schemes and survival prolongation upon early detection of the disease.

## 1. Introduction

Lung cancer remains one of the leading causes of cancer-related deaths worldwide, accounting for over 2.20 million newly diagnosed patients and over 1.7 million deaths in 2020 [1]. It is a heterogenous disease; however, over 85 % of all lung cancers are classified as non-small cell lung cancer (NSCLC), with a 5-year all-NSCLC survival rate of 28 % [1,2]. NSCLC presents with various histology, including non-squamous adenocarcinoma (NSQ) as the most prevalent form, followed by squamous cell carcinoma (SQ) [1]. The incidence of NSCLC, especially SQ, has decreased with profound smoking cessation [1,3]. Attempts to enforce regular screening for high-risk groups, even though proven beneficial, have not been fully successful yet, while overall survival (OS) has improved with the introduction of targeted and immune therapies [1,4–7].

NSCLC mutations have been widely investigated, as potential markers in the development of innovative targeted and immune therapies [1,8–11]. A proportion of subjects with advanced/metastatic NSQ

*Abbreviations:* AFT, Accelerated Failure Time; AIC, Akaike Information Criterion; ALK, Anaplastic Lymphoma Kinase; CI, Confidence Interval; CTLA4, Cytotoxic T-Lymphocyte-Associated protein 4 receptor; EGFR, Epidermal Growth Factor Receptor; ESMO, European Society for Medical Oncology; IPD, Individual Patient Data (in survival analysis); KM, Kaplan-Meier; LYG, Life Years Gained; NSQ, Non-Squamous cancer histology type; NSCLC, Non-Small Cell Lung Cancer; OS, Overall Survival; PD-1, Programmed Death receptor; PD-L1, Programmed Death-Ligand 1; PFS, Progression-Free Survival; PH, Proportional Hazards; QALY, Quality-Adjusted Life Years; RCT, Randomized Controlled Trials; RSE, Relative Standard Error; RWE/RWD, Real-World Evidence/ Real-World Data; SQ, Squamous cancer histology type; TNM staging, primary Tumor, regional lymph Nodes, distant Metastasis staging; TTE, Time-To-Event.

* Correspondence to: DSP - South Tower, Dubai Science Park, 1318, F13, OQ3, Dubai, UAE.
*E-mail address:* kirill.zhudenkov@msdecisions.tech (K. Zhudenkov).

https://doi.org/10.1016/j.csbj.2024.09.012
Received 28 June 2024; Received in revised form 19 September 2024; Accepted 20 September 2024
Available online 21 September 2024

exhibit mutations in exons 18 – 21 of tyrosine kinase domains in the epidermal growth factor receptor (exon 19 deletion or L858R EGFR mutations). EGFR mutations occur in 10–20 % of Caucasian patients with NSCLC [9,12,13]. First- and second-generation anti-EGFR treatments such as erlotinib, gefitinib and afatinib improve OS and progression-free survival (PFS) in the first-line setting [13]. Also, osimertinib, a novel selective targeted therapy for a T790M point mutation, has shown efficacy in first line for patients with various EGFR mutations [12,14].

Fusions of the echinoderm microtubule-associated protein-like 4 (EML4) gene and the anaplastic lymphoma kinase (ALK) gene translocations represent further biomarkers with corresponding developments of targeted therapies. These gene rearrangements are observed in approximately 5 % of NSCLC subjects and lead to dysregulated over-expression of ALK [15,16]. These mutations rarely co-occur with EGFR or KRAS mutations [15]. The first-generation ALK inhibitor crizotinib is a common choice as first-line therapy, although second-generation alectinib and ceritinib show improved progression-free survival [10,11,17,18]. There are other rare mutations and associated targeted treatments not discussed here including RET, NTRK, ROS1, BRAFV600E, METex14 [1,10,11].

For patients with advanced/metastatic NSCLC without driving mutations, the last decade has also been marked by critical progress linked to substantially improved survival. In particular, the development of specific antibodies against the programmed cell death protein (PD-1), its corresponding ligand 1 (PD-L1) and the cytotoxic T-lymphocyte-associated protein 4 (CTLA4) in first- or second-line setting have led to prolonged survival for a proportion of these patients [1,10,11,18]. Immune checkpoint inhibitors with anti-PD-1 and PD-L1, initially used as monotherapies in the second-line setting, have shifted toward various combination approaches. Today, with the exception of targetable mutations, a substantial number of patients with metastatic NSCLC receive an anti PD-1 or anti PD-L1 combination therapy in the first-line setting [19].

Prolonging patient survival remains one major driver in developing innovative treatments and their subsequent integration into clinical guidelines [20,21]. Key primary clinical endpoints in oncology are OS and PFS. These data represent time-to-event (TTE) outcome measures and are typically explored using Kaplan-Meier (KM) curves, which may be further stratified according to patient, biomarker, or treatment types. The KM approach, however, only represents, graphically, basic descriptive analyses and a clinical researcher cannot generate any extrapolation or simulations of alternative scenarios as well as perform a multivariate analysis.

To overcome these limitations an approach to the regression modeling of TTE data is used to qualify semi-parametric (i.e., Cox models) or fully parametric (i.e., using accelerated failure time (AFT) or proportional hazards (PH) parameterization) survival models against the study outcomes data [22]. In these models, the survival function can be estimated and then simulated and plotted over the observed TTE data, to check the goodness-of-fit [23]. Weibull models with AFT or PH parameterization has become quite popular in demographics, epidemiology, and clinical study outcome analyses [22]. The flexible parametric survival models have been proposed to model complex distributions of survival times with spline functions. The latter item is the official alias for Royston-Parmar survival models we use here [24,25]. Other approaches to the analyses of survival data have been published, including neural networks and disease progression models which make use of Markov chains [26–28].

Likewise, pharmacoeconomic analyses have been performed to assess the impact of cancer screening efficacy, where staging and progression of lung cancer as well as preventative strategies are being accounted for [4–7,26]. These modeling approaches present multiple ways to assess survival, though they are often limited to a particular sub-group of patients. There is, therefore, a lack of an integrative modeling methodology aiming at performing meta-analyses of TTE data pooled from various sources and studies, to then perform model-based indirect comparisons of various medical interventions, to subsequently predict outcomes in mixed patient cohorts which would have not yet been tested head-to head in standalone clinical trials.

The aim of this modeling research was to investigate patient survival, together with corresponding disease burden, of mixed NSCLC cohorts, through the integration of summary-level data from various sources that include randomized controlled trials (RCT) and epidemiology studies. Towards this goal, we developed a model-based platform which incorporates data on cancer stages, treatments, and survival for various patient cohorts and according to global regulatory guidelines. Here, we estimated the benefit of integrating data on multiple therapies against advanced and metastatic cancer and on disease stage distribution, to predict OS as well as to quantify disease burden in terms of life-years gained (LYG) and quality-adjusted life-years (QALY) gained.

## 2. Materials & methods

### 2.1. Data aggregation

An applied four-stage workflow for a model-based NSCLC burden assessment is shown in Fig. 1: (1) reconstruction of the treatment scheme; (2) data acquisition; (3) model development; and (4) scenario simulations. A bottom-up approach was implemented in gathering relevant data sources. First, ESMO NSCLC treatment guidelines featuring available and recommended treatments for different disease stages based on the TNM staging system were considered, to allocate patient subgroups differentiated by biomarkers, histology and treatment specifications [29,30]. Based on this analysis, a patient treatment scheme was developed, with each sub-group eligible to receive one or more approved therapies. Published RCT outcomes referenced in the ESMO guidelines were also retrieved. For early-stage cancer, where no sub-group specification was provided in the guidelines, epidemiology studies were identified to retrieve treatment details and OS for particular TNM stages [30].

Next, KM curves from the identified RCTs were digitized, to reconstruct individual patient-level datasets (IPD) using a Guyot et al. approach [31–35]. This methodology finds numerical solutions to the inverted KM equations and calculates the number of deaths at each interval from the digitized survival probability data, using the initial number of patients in a cohort. The difference between the initial number of patients and the number of deaths is the number of censoring events added to the IPD. The final survival dataset consists of IPD sub-datasets, each corresponding to either a particular treatment and/or histology or TNM stage depending on the source data. No further covariate stratification was added to the data.

### 2.2. Modeling methods

The acquired IPD were used to qualify a set of flexible Royston-Parmar survival models [36]. The best model was chosen using multiple criteria including parameter uncertainty assessment, likelihood maximization and goodness-of-fit plot analyses. The formulation of flexible parametric survival models that fit the general baseline log cumulative hazard function on the log timescale as a restricted cubic spline was applied. The log cumulative hazard function can be viewed as a linear function of log time (Eq.1) [24,25]:

$$\log H(t) = \log(\lambda) + \quad \gamma_1 = \quad \gamma_0 + \gamma_1 \log(t), \tag{1}$$

where $\gamma_0$ is a constant and $\gamma_1 \log(t)$ represents the time dependency.

When $\log H(t)$ is approximated with cubic splines, a complex survival curve shape can be described. Since natural splines are piecewise cubic polynomials, they are defined to be continuous, with continuous first and second derivatives at the knots, constrained to be linear beyond boundary knots $k_{min}$ (time of observation start), $k_{max}$ (censoring time). A

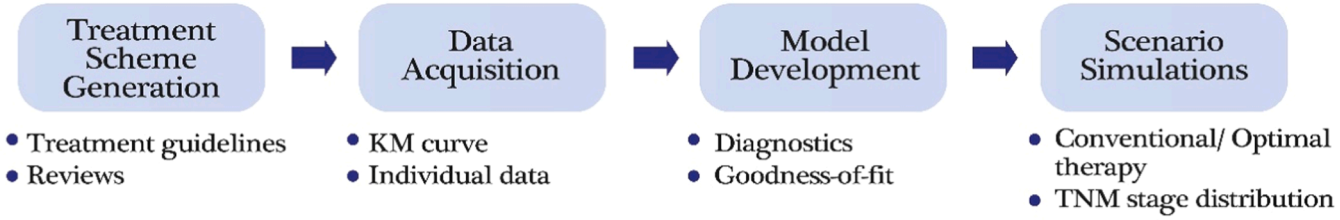


**Fig. 1.** Modeling workflow for NSCLC burden assessment.

restricted cubic spline function of log(t) can be written as (Eqs. 2–6) [25]:

$$s(\log(t)) = \gamma_0 + \gamma_1 z_1 + \gamma_2 z_1 + \ldots + y_{m+1} z_m, \quad (2)$$

where $z_j$ is the $j^{th}$ basis function

$$z_1 = \log(t) \quad (3)$$

$$z_j = (\log(t) - k_j)_+^3 - \lambda_j(\log(t) - k_{min})_+^3 - (1-\lambda_j)(\log(t) - k_{max})_+^3, \quad (4)$$

$$(\log(t) - k_a)_+^3 = \begin{cases} (\log(t) - k_a)^3, \text{if} \log(t) > k_a \\ 0, \text{otherwise} \end{cases} \quad (5)$$

$$\lambda_j = \frac{k_{max} - k_j}{k_{max} - k_{min}}, \quad (6)$$

where $k_j$ is $j^{th}$ knot position, $k_a$ is any knot position. If $m = 0$, then there are only two parameters, $\gamma_0$, $\gamma_1$, defining a Weibull survival model [36].

Optimal model development including knot position selection was performed based on the lowest Akaike information criterion (AIC) and RSE less than 50 % for all model parameters. The qualified γ parameters and knot positions alone can be used to construct the mean probability survival trend; however, here we also incorporate survival function uncertainty by sampling 500 sets of model parameters from the multivariate normal distribution, based on the obtained variance-covariance matrix [37,38]. The 2.5 % and 97.5 % percentiles of the generated pool of survival functions' estimates were then derived to construct 95 % CI. The time interval for simulations was up to 60 months for the full NSCLC patient cohort and up to 24 months for the stage IIIb/IV sub-cohort (advanced/metastatic NSCLC).

When simulating survival of the full patient cohort or for the multiple treatments incorporated in stage IIIb/IV, the summary function describing OS in respective cohort was defined as a mixture of survival models corresponding to particular sub-groups, with a weighting parameter reflecting the expected proportion of these sub-groups within the whole patient cohort; OS was, therefore, calculated as a sum of continuously sampled estimates (Eq.7):

$$s_{full}(\log(t)) = a_1 s_1(\log(t)) + a_2 s_2(\log(t)) + \ldots + a_n s_n(\log(t)) \quad (7)$$

where $a_n$ is the proportion of an n sub-cohort in the full population, $s_n$ is the survival function describing survival for the $n^{th}$ sub-cohort.

### 2.3. Disease burden evaluation

Patient quality of life (QoL) was integrated to estimate the disease burden weighed on utility scores. For each pairwise comparison of scenarios, the LYG was calculated as a difference in area under the two respective survival curves (AUC) (Eqs. 8,9,10); then weighted on utility scores for the corresponding stages of NSCLC, to derive QALY gained (Eq. 11) [4,39]. For this analysis, we allocated utility scores depending on the cancer stage alone, much like in a typical pharmacoeconomic analysis [4,5,7,26].

$$AUC_1 = \int_0^T a_{11} * s_{11}(\log(t)) + \int_0^T a_{12} * s_{12}(\log(t)) + \ldots + \int_0^T a_{1n} * s_{1n}(\log(t)) \quad (8)$$

$$AUC_2 = \int_0^T a_{21} * s_{21}(\log(t)) + \int_0^T a_{22} * s_{22}(\log(t)) + \ldots + \int_0^T a_{2n} * s_{2n}(\log(t)) \quad (9)$$

$$LYG = AUC_2 - AUC_1 \quad (10)$$

$$QALY = u_1 * (a_{21}\int_0^T s_{21}(\log(t)) - a_{11}\int_0^T s_{11}(\log(t))) + u_2 * (a_{22}\int_0^T s_{22}(\log(t)) - a_{12}\int_0^T s_{12}(\log(t))) + \ldots + u_n * \left(a_{2n}\int_0^T s_{2n}(\log(t)) - a_{1n}\int_0^T s_{1n}(\log(t))\right), \quad (11)$$

where $u_n$ is a utility score associated with a particular stage and T is the righthand time cut point of interest for LYG or QALY. Values for T were set either to 24 or 60 months depending on the chosen scenario, advanced/metastatic NSCLC, or all NSCLC stages analyzed, respectively.

The presented flexible survival models and disease burden evaluation were additionally used to generate survival simulations of different patient cohorts representing optimal treatment introduction effect as well as the impact of variable initial distribution of cancer stages by means of sensitivity analyses.

### 2.4. Software

OS data digitization was performed in *Plot Digitizer* [40]. IPD reconstruction, model development, simulations and pharmacoeconomic calculations were performed in *R statistics* (version 4.2.2) [41]. *R* packages *flexsurv_2.2.2* and *survival_3.4.0* were used for model qualification. Other *R* packages used included *MASS_7.3–58.1* for parameter estimation from the multivariate distribution, *rstpm2_1.6.3* for the combination of survival functions and LYG and QALY gained

calculations, and *ggplot2_3.4.1* for OS visualization.

Web-based version of the modeling platform is deployed at ShinyApps repository and can be accessed at https://www.oncomonitor.tech (see more details in the Fig. S5, address the corresponding author for more information if needed).

## 3. Results

### 3.1. Model development

First, ESMO guidelines for NSCLC were examined to determine cancer staging and major types of recommended treatments available on the market. Overall, a whole NSCLC cohort was composed of four sub-cohorts based on the TNM classification [30]. Since early stages of NSCLC are treated mostly surgically with adjuvant therapy, it was decided to use an epidemiology source to retrieve OS. KM curves describing early NSCLC patient survival who underwent surgery and adjuvant treatment (13469 patients total) were digitized [30]. Aggregated OS data from sub-stages were then transferred into the IPD and combined into three IPD datasets for stages I, II, and IIIa, representing typical survival in Caucasian population.

Next, representative classes of first-line treatments were established for advanced and metastatic NSCLC (stages IIIb/IV) from RCT outcomes of the analyses referenced in ESMO guidlines [29]. A broader range of recommended treatments allowed for pooling multiple classes of treatments representing both a conventional standard of care and relatively novel/optimal therapies. Variability in treatments was limited to targeted therapies (ALK inhibitor therapy and anti-EGFR treatment), immunotherapies (PD-1 inhibitors), and chemotherapies (platinum-based chemotherapy doublets). To eliminate a potential outcome bias, clinical studies for Asian only or elderly patients only were not included, since it is known that these patients exhibit a significantly different survival when compared to non-Asian or non-elderly groups [42–44]. Likewise, clinical studies including patients with poor performance status and heavy smokers were excluded, for the same reason [45]. For respective studies the long-term outcomes were checked for being published and collected for analysis.

As a result, a total of ten KM curves were selected to describe survival of the subjects with advanced NSCLC using the RCT outcome data (2339 patients total) [46–50]. This information was utilized to reconstruct IPD and create two datasets describing conventional standard of care (alectinib/crizotinib, gefitinib/erlotinib, chemotherapy) or optimal treatment (alectinib/crizotinib, osimertinib, pembrolizumab mono- or combination therapy) for patients with ALK-translocation, EGFR mutation and without driver mutations, respectively. Based on the retrieved data, a generalized treatment scheme with proportional patient distribution based on stages, histology and relevant biomarkers was created to describe OS and respective disease burden (Fig. 2). More details on the data used are available in Table S6.

When the complete IPD dataset was created, a set of models were qualified against the treatment data, to produce a comprehensive survival model for the full NSCLC cohort. The tested models included parameterization with cubic splines with the number of knots ranging from 0 to 2 (Table S1) [24,25]. A typical survival curve for each treatment was parameterized with gamma parameters and knot position values (Table S2a) according to the cubic spline formulae (Eqs. 2–6) [24, 25].

Uncertainty of survival function was derived as explained in Materials and Methods (see the parameter estimates and variance-covariance matrices in Table S2b). The output from this step resulted in 12 survival models, each describing a treatment scheme as shown in Fig. 2. The optimal models had the lowest AIC criterion values yet keeping parameter identifiability (RSE<50 %, see Table S1). The goodness-of-fit plots (Fig. S1, S2) showed adequate reproduction of KM data.

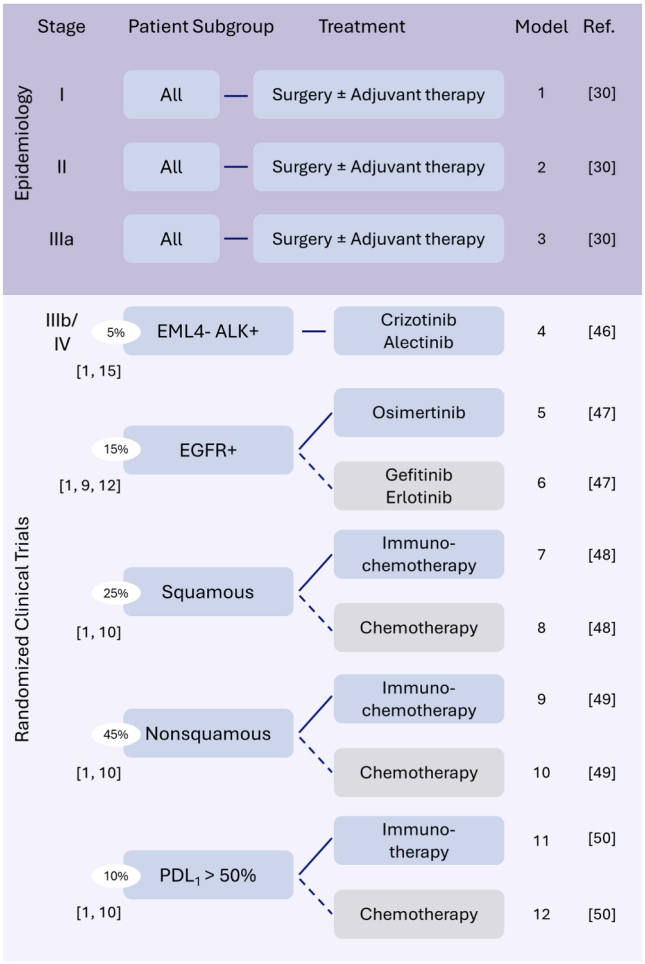


**Fig. 2.** NSCLC treatment scheme with different survival models. Conventional treatment is shown in grey, optimal treatment is shown in blue. Prevalence of each sub-group in stage IIIb/IV is shown as a percentage in an oval. References are provided in square brackets.

### 3.2. Survival simulations for advanced and metastatic NSCLC patients

The family of survival models we developed was used to predict and compare OS across a variety of scenarios, including specialized targeted and immune treatments that have been tested and approved over the last decade (Fig. 2). Due to the presence of various drug classes, analyses were performed using one or two representative treatments from the same class. First, a comparison among the same treatment types was performed, to visualize the effect in patient survival with differing histology (SQ, NSQ) and driver mutations (ALK or EGFR). The resulting simulations show that 95 % confidence intervals intersect among survival curves for the same treatment class in the EGFR driving mutation pool (Fig. S3a) and for the same treatment for different histology (Fig. S3b-c).

Then the evaluation of the incremental benefit of specialized/optimal treatments versus the conventional standard of care in the advanced NSCLC patient sub-group was performed (Fig. 3). OS was analyzed in four distinct patient sub-populations: NSQ cancer without a driver mutation, SQ cancer without a driver mutation, NSQ or SQ cancer with high PD-L1 expression (>50 %) [29,51], and NSQ cancer with an EGFR+ driver mutation. A comparative simulation for the ALK+ patient sub-group was not introduced since there was quite minor difference detected on the levels the complete IIIb/IV patient subgroup and whole cohort outcomes (up to 0.1 months difference in median survival; data not shown). Despite the difference between OS for crizotinib and

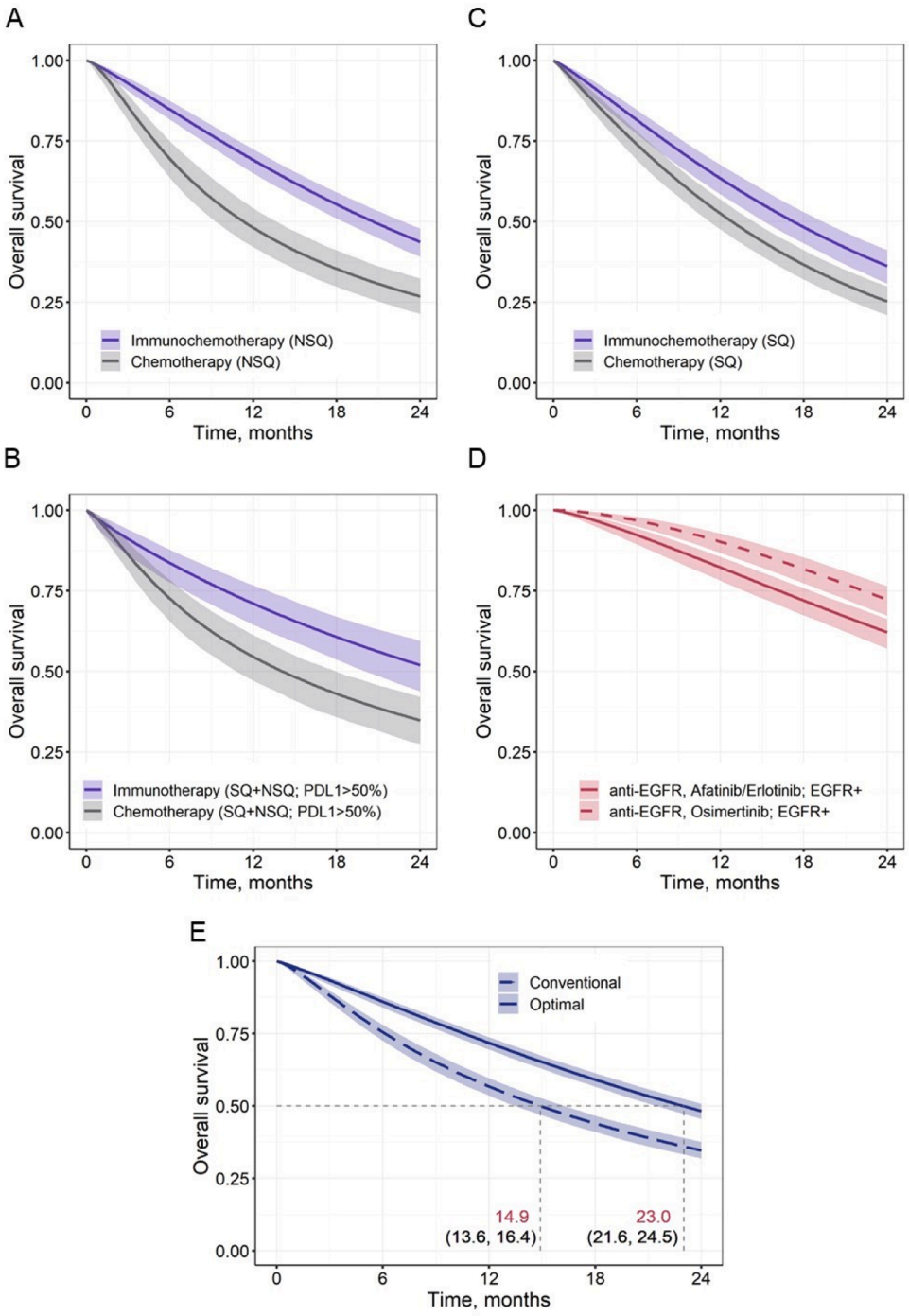


**Fig. 3.** Overall survival comparison for optimal vs. conventional therapy for selected sub-cohorts with advanced NSCLC. A. NSQ histology: chemotherapy, immunochemotherapy. B. SQ histology: chemotherapy, immunochemotherapy. C. NSQ/SQ histology: chemotherapy, immunochemotherapy. D. EGFR+ : anti-EGFR targeted therapy. The bands indicate 95 %CI for model simulations. E. All patients: all treatments.

alectinib [46], the prevalence of ALK+ subjects is quite small, thus for the simulations the mixed alectinib/crizotinib outcomes were taken to represent OS in all scenarios.

Conventional treatment schemes were taken as follows: alectinib/crizotinib for the ALK+ sub-cohort, gefitinib/erlotinib for the EGFR+ sub-cohort, and chemotherapy for both SQ or NSQ sub-cohorts with or without high PD-L1 expression. The optimal treatment incorporated alectinib/crizotinib for the ALK+ sub-cohort, osimertinib for the EGFR+ sub-cohort, pembrolizumab monotherapy for SQ/NSQ sub-cohorts with high PD-L1 expression and pembrolizumab combination with chemotherapy for SQ and NSQ sub-cohorts without any driver mutations. For NSQ and SQ sub-cohorts not bearing any driver mutations and with high PD-L1 expression, a prolongation in OS was observed when immunochemotherapy was administered, as compared to a conventional platinum doublet chemotherapy (Fig. 3a-c). An optimized treatment with osimertinib as an anti-EGFR treatment showed an improvement versus conventional anti-EGFR therapy (Fig. 3d).

Simulations showed that there was a significant difference between optimal and conventional treatments and the 95 %CI did not overlap. Conventional vs. optimal treatment scenarios in a cohort with advanced/metastatic NSCLC with all treatment types incorporated showed median survival time for the conventional treatment option

**Table 1**
Stage distribution, LYG and QALY for selected simulation scenarios.

| All Stages (up to 60 months) | Test/Baseline scenario | Stage distribution (%) | | | | LYG (95 %CI), months | QALY gained (95 %CI), months |
|---|---|---|---|---|---|---|---|
| | | I | II | IIIa | IIIb/IV | | |
| | Utility Score | 0.87 | 0.87 | 0.77 | 0.57 | | |
| | Scenario 2: Optimal | 35 | 15 | 10 | 40 | 2.27 (1.47,3.00) | 1.30 (0.84,1.71) |
| | Scenario 2: Conventional | 35 | 15 | 10 | 40 | - | - |
| | Scenario 1: Optimal | 29 | 8 | 13 | 50 | 2.84 (1.83,3.74) | 1.62 (1.04,2.13) |
| | Scenario 1: Conventional | 29 | 8 | 13 | 50 | - | - |
| | Standard: Optimal | 15 | 14 | 14 | 57 | 3.24 (2.09,4.27) | 1.85 (1.19,2.43) |
| | Standard: Conventional | 15 | 14 | 14 | 57 | - | - |
| | Scenario 2: Optimal | 35 | 15 | 10 | 40 | 8.40 (7.45,9.31) | 7.86 (7.31,8.38) |
| | Standard: Conventional | 15 | 14 | 14 | 57 | - | - |
| Stage IIIb/IV patients (up to 24 months) | Optimal | 0 | 0 | 0 | 100 | 2.89 (2.15, 3.58) | 1.65 (1.22 2.04) |
| | Conventional | 0 | 0 | 0 | 100 | - | - |

equal to 14.9 months (95 % CI: 13.6, 16.4) and for the optimal treatment – 23.0 months (95 % CI: 21.6, 24.5) (Fig. 3e). Therefore, a cumulative benefit from the introduction of optimal set of therapies was 8.1 months (95 % CI: 5.9, 10.0) (Table S3).

### 3.3. Survival simulations for the full NSCLC patient cohort

To illustrate the sensitivity of the model-based OS predictions to cancer stage prevalence in the cohort, a range of simulations were performed reflecting actual NSCLC populations and advanced cancer treatment schemes as described above (conventional vs. optimal). Three stage distributions at the time of diagnosis (representing different cancer screening efficacy) were tested to evaluate OS up to 60 months (Table 1). OS was simulated for a standard stage prevalence ("Standard") [52], where stages I, II, IIIa and IIIb/IV were distributed as follows: 15 %, 14 %, 14 %, and 57 %. Also, two scenarios with higher prevalence of early cancer stages representing more effective detection ("Scenario 1″ and "Scenario 2″) [53,54] were considered, with distributions of, respectively, 29 %, 8 %, 13 %, 50 % (Scenario 1) and 35 %, 15 %, 10 %, 40 % (Scenario 2). These stage distribution samples were published recently for Russia [52], the USA [53] and EU [54].

Simulations for conventional treatment showed an improvement in OS for Scenario 1 and Scenario 2, when compared to Standard: with OS of 32.4 months (95 % CI: 30.5, 34.8) for Scenario 1, 43.2 months (95 % CI: 41.1, 45.9) for Scenario 2 and 25.6 months (95 % CI: 24.1, 27.4) for Standard scenario, i.e., a gain in OS over Standard scenario of 6.9 months (95 % CI 6.3, 7.6) for Scenario 1 and 17.6 months (95 % CI: 16.5, 19.0) for Scenario 2, as shown in Fig. S4a and Table S3. For an optimal treatment, the OS was 32.4 months (95 % CI: 30.8, 34.2) for Standard scenario, 39.1 months (95 % CI: 37.1, 41.2) for Scenario 1, and 48.1 months (95 % CI: 46.0, 50.4) for Scenario 2. Therefore, OS increased by 6.66 months (95 % CI: 6.2, 7.2) and 15.7 months (95 % CI: 14.8, 16.7), for Scenario 1 and Scenario 2 respectively (Fig. S4b, Table S3).

Pairwise comparisons were also tested, where stage distribution was kept constant and the effect of treatment for stage IIIb/IV disease was evaluated, in addition to the effect of optimal treatment and prevalent early-stage distribution at detection (Fig. 4). Simulations for optimal treatment showed an improvement in OS of 32.4 months (95 % CI: 30.7, 34.1) for Standard scenario, 39.1 months (95 % CI: 37.1, 41.2) for Scenario 1, 48.1 months (95 % CI: 46.1, 50.4) for Scenario 2 in Table 4a-c. Therefore, an introduction of optimal therapies shows a gain in OS over conventional treatment is 4.9 months (95 % CI: 1.9, 7.6) for Standard Scenario, 6.7 months (95 % CI: 3.8, 9.4) for Scenario 1, and 6.9 months (95 % CI: 4.6, 9.1) for Scenario 2 in Table S3.

We can observe the profound effect of optimal therapies while we are investigating the advanced NSCLC subgroup only (Fig. 3). These simulations show that in terms of OS benefit, the most efficient is Scenario 2 with optimal treatment introduced (Fig. 4c-d). Also, in a longer-term perspective we see a tendency of elimination of the effect: despite of quite high initial prevalence, patients with advanced NSCLC have shorter survival; thus, their longer-term impact becomes limited, which is examined further in the sensitivity analysis.

### 3.4. Disease burden evaluation

A further effect assessment of optimal and conventional treatments as well as cancer stage prevalence on disease burden was conducted by estimating standardized pharmacoeconomic metrics such as LYG and QALY gained. Given the survival probability derived from the simulations described above, we calculated LYG estimates for all pairwise scenario combinations; we also computed QALY gained, which corresponds to LYG weighted with utility scores (0.87 for stage I, 0.87 – stage II, 0.77 – stage IIIa and 0.57 – stages IIIb/IV) for the time horizon of 60 months (24 months for stages IIIb/IV only) [4] (Table 1 & S4).

When considering all stages and depending on the scenario, the introduction of optimal therapies provided LYG values ranging from 2.3 (Scenario 2) to 3.2 months (Standard scenario) and QALY gained scores from 1.3 (Scenario 2) to 1.9 months (Standard scenario) (Table 1). The pharmacoeconomic metrics for Scenario 1 were in-between those for Standard scenario and Scenario 2. The effect of a higher early NSCLC prevalence (representing more efficient screening), in terms of LYG, ranged from 2.6 to 3.0 months (Scenario 1), and from 5.2 to 6.1 months (Scenario 2), where larger values represent cohorts on conventional treatment. The corresponding QALY gained values of 2.9 to 3.1 months (Scenario 1) and 6.0 to 6.6 months (Scenario 2) showed the same tendency favoring conventional treatment. Likewise, when the cumulative effect of optimal treatment and earlier disease stage diagnosis prevalence (Scenario 2) was estimated, LYG and QALY gained were, respectively, 8.4 and 7.9 months.

These results suggest that the introduction of optimal therapies alone yielded a smaller improvement, for both LYG and QALY gained, in the all-stages cohort in contrast to a more efficient detection with resultant earlier NSCLC detection represented by Scenarios 1 and 2.

When looking at advanced NSCLC population only, then optimal therapies provide values of 2.9 months of LYG and 1.7 months of QALY gained. These values, however, while proving a positive effect of an optimal treatment, do not infer the magnitude of the effect when the percentage of stages IIIb/IV cases is changed.

To estimate the effect that advanced/metastatic NSCLC (stage IIIb/IV) prevalence may have on disease burden for the whole cohort, a sensitivity analysis was performed. The main goal was to determine the dynamics of LYG and QALY as the prevalence of stages IIIb/IV varies from 30 to 60 % with a set of scenarios based on the Standard scenario distribution (Table S5). Since stages I, II an IIIa vary as well in this scenario, two strategies were examined: (1) Stages I, II and IIIa are equally distributed, (2) Stage I is always more prevalent than Stage II and Stage IIIa, which are kept constant at 13 % of the whole population. Thus, sampling was performed at 30 %, 35 %, 40 %, 45 %, 50 %, 55 % and 60 % for Stage IIIb/IV prevalence for both options (Table S5). To

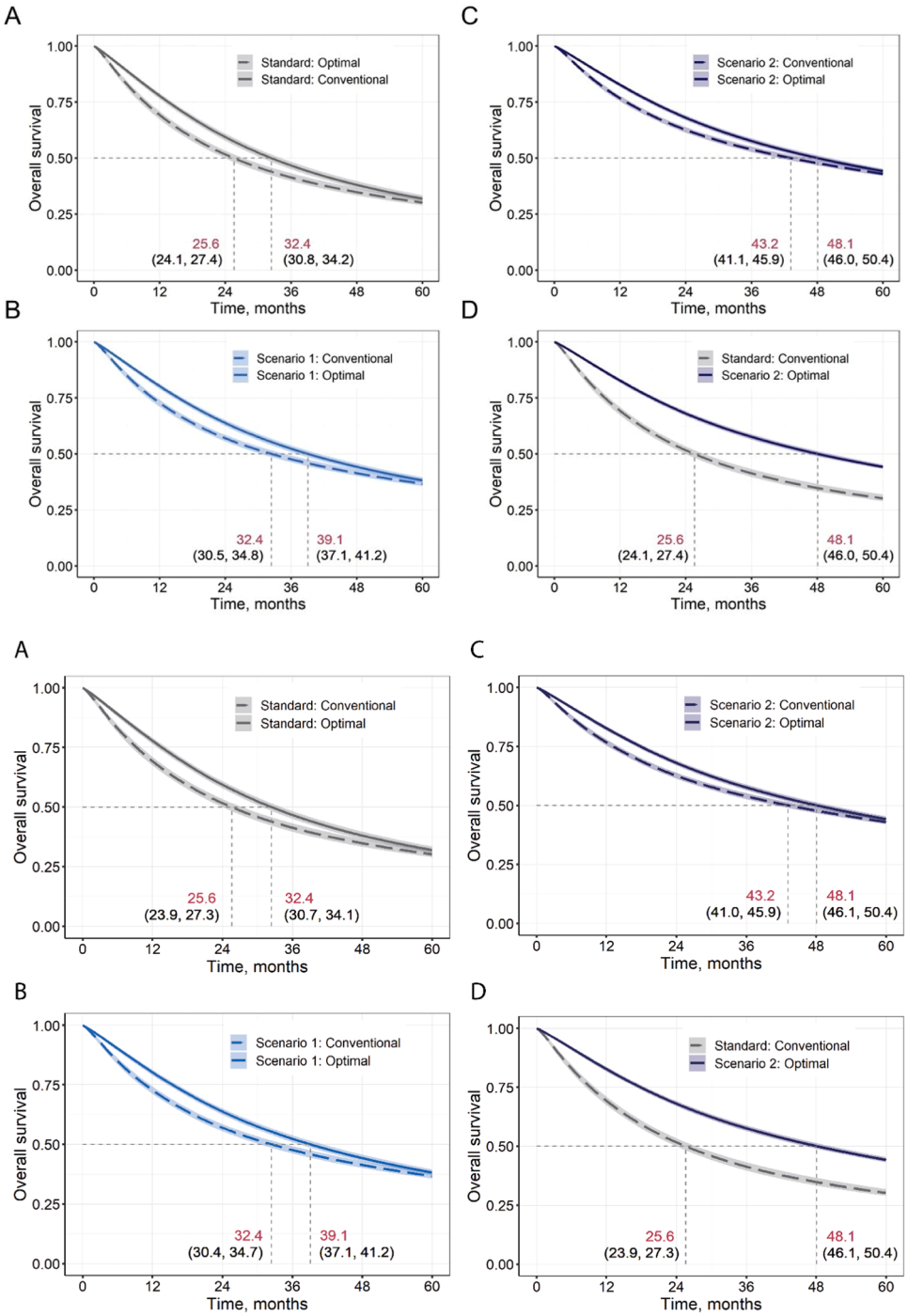


**Fig. 4.** Overall survival comparisons obtained for different scenarios of cancer stage distribution and optimal therapy introduction. Panels represent OS and its 95 % CI for different scenarios. A – C. Optimal and conventional (solid and dashed lines) treatments applied for advanced NSCLC effect for the whole cohort (Standard scenario, Scenario 1, and Scenario 2, respectively). D. Scenario 2 with optimal therapy compared to the Standard Scenario with conventional therapy. Numbers within graphs represent median survival (red) and its 95 % CI (black).

collect LYG and QALY gained estimates, for all tested scenarios a comparison to the Standard Scenario (Conventional or Optimal therapy) was performed.

The sensitivity analysis revealed an overall increase in LYG and QALY as the prevalence of stages IIIb/IV decreased from the baseline 60 % (Fig. 5). In terms of two strategies, most beneficial for LYG and QALY, was the case, where stage I is prevalent, representing the most efficient screening positively impacting both survival and quality of life for patients.

However, it should be noted that in these scenarios we observe that the conventional treatment comparison provided higher LYG and QALY gained than the optimal treatment comparison. It is reasonable due to the higher efficacy of optimal treatment for Stage IIIb/IV – the more efficiently we treat cancer at late stages, the less effect will be retrieved while increasing early stages prevalence in the whole cohort.

In summary, when the NSCLC patient population predominantly consists of the subjects with advanced cancer, the introduction of optimal therapies results in a smaller increase in LYG and QALY gained, in comparison to scenarios with more efficient NSCLC detection and predominant early NSCLC, where an overall incremental benefit of an

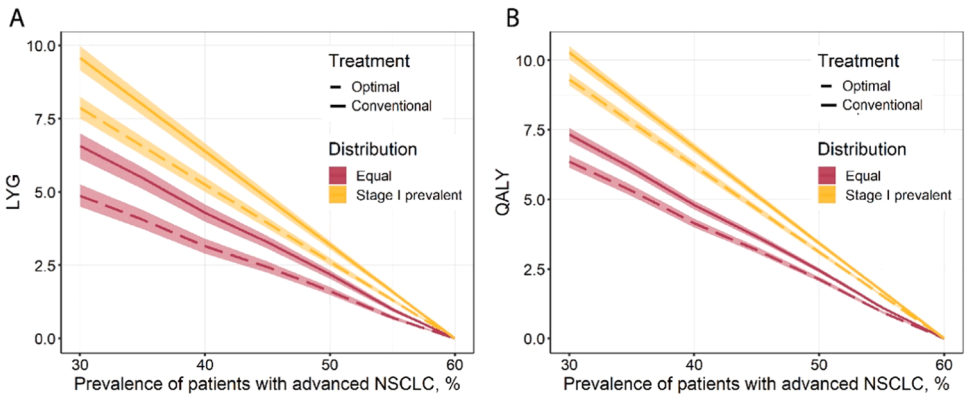


**Fig. 5.** Sensitivity analysis of advanced NSCLC prevalence. Effect of changing prevalence of advanced stages on NSCLC disease burden. Bands indicate 95 % confidence intervals. Line type represents different treatment schemes for advanced stages (dashed – optimal therapy, solid – conventional therapy). Colors represent distribution options (red – equal distribution of stages I, II, IIIa, yellow – stages II & IIIa are kept constant, stage I is prevalent). A. LYG. B. QALY gained.

optimal treatment would become more prominent, especially if stage I is more prevalent in the patient population.

## 4. Discussion

Predictive mathematical models applied to the assessment of patient survival in a particular cancer such as NSCLC can be a powerful approach in the evaluation of disease burden for a given set of therapies and in a heterogeneous patient population.

Typically, RCT outcomes in oncology are reported as OS and PFS and the non-parametric data analytics applied are limited to basic descriptive statistics. More advanced analyses are required to provide a comprehensive tool to effectively predict OS. Multiple approaches based on neural networks [27], Markov models [26,55,56] and meta-analyses [44,57,58] have been proposed, yet they are not fit to solve the problem highlighted in this research. While being popular and relying on large datasets, machine learning approaches lack the deterministic basis that is required for predictive simulations/extrapolations (i.e., for scenarios not included into the calibration dataset) [27,59]. Markov models, on the contrary, are traditionally used in decision-making modeling in health outcome studies [26,54,60,61]. These models have proven utility in certain settings; however, they cannot be used for extrapolations to various treatments and populations if individual data for the calibration of particular Markov chain transitions are limited [26,55,60].

Likewise, pharmacoeconomic and cost-effectiveness analyses have been performed to assess the benefits of more efficient cancer screening, where cancer staging and progression as well as preventive strategies are being accounted for [4–7,26]. Even though these model-based analyses represent various approaches to survival assessment, they usually are limited to a specific sub-group of patients. When it comes to the analysis of survival data of the whole patient cohort, it becomes a challenge to integrate multiple sources of data, due to a high level of heterogeneity in studies and patient populations.

One may operate with point estimates, such as median OS and PFS, including 1-,2-, 5-year milestone survival characteristics, to build out a standard meta-analysis [57,62]. Such an approach, however, provides only the metrics that were included in the analysis – for example, median survival cannot be retrieved, if standard meta-analysis is performed for another metric, long-term 5-year OS. There is, therefore, a lack of methodologies to perform meta-analyses of exhaustive TTE data pooled from numerous studies and run model-based indirect comparisons of various treatments in whole patient cohorts which would not have been tested head-to head in standalone RCTs. In the present study, a comprehensive model-based approach to assess NSCLC survival and burden was developed.

The treatment effect in stratified patient sub-groups representing different cancer stages, treatments, and targetable mutations was described using flexible AFT survival models. Subsequently, OS in the whole patient cohort was calculated as a weighted sum of survival functions corresponding to clinically relevant proportions between specific sub-groups of patients with early and advanced NSCLC.

A key novelty of our modeling approach lies in the ability of the model to produce personalized OS predictions, for patient cohorts presenting a variety of baseline characteristics. Single RCTs focus on the outcomes in dedicated patient cohorts and following a set of inclusion criteria; thus, it would not be possible to determine the outcomes for the whole cohort. On the other hand, epidemiology data represent a good source of pooled OS data, yet these data are retrospective and take decades to be compiled and published; furthermore, they might not include outcomes data in specific patient sub-groups of interest. Here, we combine both RCT and epidemiology data sources, to provide a robust modeling approach to predict long-term OS.

We incorporated the effects of optimal therapies from single RCTs in advanced and metastatic NSCLC, including checkpoint inhibitor immunotherapies and targeted treatments. We evaluated how a substantial benefit of a single therapy in a selected sub-population (advanced NSCLC) translates into a lesser benefit in the whole patient population. For example, in the osimertinib RCT [49], an astounding effect in OS prolongation was shown, as compared to standard anti-EGFR therapies, yet only ~15 % of the NSCLC population feature a targetable EGFR mutation and, therefore, the magnitude of benefits that osimertinib shows in the RCT will be smaller in a heterogeneous cohort in RWE/RWD setting. However, in Asian population the prevalence of EGFR mutation may be remarkably higher (up to 48 %) [12,63]. So that may accordingly prolong the survival of the whole cohort (Fig. S5 representing the graphical user interface of the platform running a special scenario with higher prevalence of EGFR subjects).

Finally, in current analysis for Caucasian population, osimertinib would not be expected to have the same impact in a heterogeneous NSCLC population as compared to, for example, checkpoint inhibitor immunotherapies, which recently have been tested in patients with or without high PD-L1 expression and show improvements in all eligible patient sub-groups, accounting for over 80 % of all NSCLC patients [1,19]. Here we consider pembrolizumab a good representation of immune therapy as it is explicitly mentioned in the guides [64]. However, an additional meta-analysis of immune checkpoint inhibitors therapy comparative efficacy may be further introduced using the methods presented in this research or utilizing the other techniques [65]. The modeling platform proposed here may allow, in fact, to quantify an efficacy target for the whole NSCLC population, upon considerations of treatment effects in specific sub-groups of patients, prevalence of these patients in the whole population, as well as other relevant factors.

In the present work, we evaluated improvements in OS for the whole NSCLC cohort, if an increased fraction of patients with early NSCLC were

to be identified, e.g., based upon improved detection / detection procedures. To further understand the effect of therapies in a whole cohort of patients, we inspected how early-stage NSCLC prevalence contributes to OS. Our modeling findings together with published sources suggest that it is more effective to invest into detection procedures than to source costly specialized treatments, since earlier cancer detection leads to prolonged OS [1,26,66]. A similar result was previously obtained using Markov Chain modeling based on a cost-effectiveness analysis of various detection methods, supporting the use of more frequent (e.g., bi-annual) cancer detection [26].

Since it has been established that higher early-stage NSCLC prevalence plays an important role in improving survival in the mixed cohorts, the presented modeling platform can be extended to include innovative therapies for earlier disease stages, to augment the number of therapies derived from different sources [30]. For example, a novel treatment such as pembrolizumab or durvalumab in unresectable cancer can be introduced, to quantify further survival prolongation of earlier-stage patients (for the subjects with stage III, locally advanced NSCLC) [67–70]. However, for these innovative treatment schemes OS is available only for previously treated subjects to the moment. Likewise, RCTs with various radiotherapy scenarios may also be integrated into the treatment scheme and parameterized accordingly, as a first-line treatment option for localized tumors [71,72]. These updates will increase the sensitivity of the outcome prediction to treatment scheme details.

The benefits brought upon by innovative therapies remain a topic of interest, from both perspectives of prolonging the life of patients and finding the most economically effective approach. Here, to further illustrate applications of our modeling platform, we presented pharmacoeconomic evaluations based on LYG and QALY gained, thereby quantifying the benefit of earlier disease detection, with optimal treatment schemes. Our findings are supported by multiple cost-effective analyses evaluating various detection programs and prevention approaches, to enforce early NSCLC detection [4–7]. Thus, both timely cancer diagnostics and optimal therapy application are drivers of prolonged OS and improved quality of life. Also, it should be noted that quantitative health economics outcomes may depend on utility assessment for different stages [4,73].

Likewise, since there are various methods to evaluate interventions and apart from their impact on survival, expansions of the presented modeling approach can be made to introduce respective costs per QALY gained for the considered scenarios representing cost-effectiveness assessments. Multiple confounding variables such as treatment availability and corresponding cost may also be researched and integrated, to predict disease burden for specific regions and countries. It is noteworthy that in current analyses the acquired LYG and QALY gained characteristics are comparable within the scenarios for the same time intervals – up to 24 months (for advanced NSCLC sub-cohort) and up to 60 months (for the whole cohort). Within the same time interval, the survival outcomes as well as LYG are comparable for differently composed (by treatment and disease stages) scenarios. QALY gained comparability is acquired by means of proper weighting on heath state utilities.

The modeling platform presented here illustrates how RCTs and epidemiology data may be used to describe survival of a whole patient cohort. The results we derived may lack the precision and predictive accuracy of a "pure" model-based meta-analysis, due to the absence of individual patient-level data [58,74] as well as external data for model validation. Additional sources of individual patient-level RWE/RWD and RCT outcome inclusion are desirable to overcome these challenges.

In conclusion, the modeling platform presented here provides an integrative tool to assess and predict survival in mixed cohorts of patients; it also allows for an assessment of quality of life and health economics characteristics. This platform also offers a robust approach to quantify the benefits of early cancer screening/detection over the whole, highly heterogeneous population. Furthermore, the effect of optimal/innovative treatment for an advanced and metastatic NSCLC sub-cohort was evaluated. The estimated survival prolongation was extended to quality-of-life adjustment, to describe cancer burden. This platform can be applied by pharmaceutical developers and drug regulators to determine necessary changes in treatment and screening regimens, to meaningfully prolong OS to targeted margins. A further potential application of this platform lies in health economics and budget impact assessments, in support of decision-making at different levels in healthcare.

## Funding

No funding.

## CRediT authorship contribution statement

**Kirill Zhudenkov:** Writing – review & editing, Writing – original draft, Supervision, Conceptualization. **Kirill Peskov:** Writing – review & editing, Writing – original draft, Supervision. **Victoria Kulesh:** Writing – review & editing, Visualization, Software. **Gabriel Helmlinger:** Writing – review & editing, Writing – original draft, Methodology. **Boris Shulgin:** Methodology. **Nikolai Katuninks:** Software, Methodology. **Nataliya Kudryashova:** Writing – original draft, Visualization, Software, Investigation, Data curation, Conceptualization.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

This research work was supported by the Academic leadership program Priority 2030 proposed by Federal State Autonomous Educational Institution of Higher Education I.M. Sechenov First Moscow State Medical University of the Ministry of Health of the Russian Federation (Sechenov University). We would also like to thank Oleg Stepanov, Evgenii Semenov, Kseniya Danichkina and Anna Mishina for the contribution they made to this paper.

### *Author agreement statement*

We the undersigned declare that this manuscript is original, has not been published before and is not currently being considered for publication elsewhere.

We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed.

We further confirm that the order of authors listed in the manuscript has been approved by all of us.

We understand that the Corresponding Author is the sole contact for the Editorial process. He/she is responsible for communicating with the other authors about progress, submissions of revisions and final approval of proofs.

### *Access to data*

Previously published data only were used for this research, using open-source database search engines.

### *Author contributions*

Nataliya Kudryashova: Investigation, Data curation, Writing - Original draft preparation, Visualization, Conceptualization, Methodology, Software; Boris Shulgin: Methodology; Nikolai Katuninks: Methodology, Software; Kirill Zhudenkov: Conceptualization, Supervision, Writing - Reviewing and Editing; Victoria Kulesh – Editing, software and online

platform deployment; Gabriel Helmlinger: Methodology, Writing – Reviewing and Editing; Kirill Peskov: Supervision, Writing - Reviewing and Editing.

*Informed consent and patient data*

The study is a retrospective analysis of aggregated data from published clinical study results, demography, and epidemiology. No study medication was supplied, and no patients were entered into this computational modeling study. There was no requirement for informed consent in this computational modeling study, which made use of previously published, cohort summary-level data only.

## Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2024.09.012.

## References


[1] Thai AA, Solomon BJ, Sequist LV, Gainor JF, Heist RS. Lung cancer. Lancet 2021; 398(10299):535–54. https://doi.org/10.1016/S0140-6736(21)00312-3.

[2] Cancer.Net Editorial Board. Lung Cancer - Non-Small Cell: Statistics. Cancer.Net ® ASCO: Knowledge Conquers Cancer. Published February 2022. Accessed June 23, 2022. ⟨https://www.cancer.net/cancer-types/lung-cancer-non-small-cell/statistics⟩.

[3] Zhang YH, Lu Y, Lu H, Zhou YM. Development of a survival prognostic model for non-small cell lung cancer. Front Oncol 2020;10:362. https://doi.org/10.3389/fonc.2020.00362.

[4] Kowada A. Cost-effectiveness and health impact of lung cancer screening with low-dose computed tomography for never smokers in Japan and the United States: a modelling study. BMC Pulm Med 2022;22(1):19. https://doi.org/10.1186/s12890-021-01805-y.

[5] Diaz M, Garcia M, Vidal C, et al. Health and economic impact at a population level of both primary and secondary preventive lung cancer interventions: a model-based cost-effectiveness analysis. Lung Cancer 2021;159:153–61. https://doi.org/10.1016/j.lungcan.2021.06.027.

[6] Van Meerbeeck JP, O'Dowd E, Ward B, Van Schil P, Snoeckx A. Lung cancer screening: new perspective and challenges in Europe. Cancers 2022;14(9):2343. https://doi.org/10.3390/cancers14092343.

[7] Veronesi G, Navone N, Novellis P, et al. Favorable incremental cost-effectiveness ratio for lung cancer screening in Italy. Lung Cancer 2020;143:73–9. https://doi.org/10.1016/j.lungcan.2020.03.015.

[8] Sun L, Marmarelis ME, Langer CJ. Systemic therapy for mutation-driven NSCLC. Semin Radiat Oncol 2021;31(2):140–8. https://doi.org/10.1016/j.semradonc.2020.11.011.

[9] Hsu WH, Yang JCH, Mok TS, Loong HH. Overview of current systemic management of EGFR-mutant NSCLC. Ann Oncol 2018;29:i3–9. https://doi.org/10.1093/annonc/mdx702.

[10] Mithoowani H, Febbraro M. Non-small-cell lung cancer in 2022: a review for general practitioners in oncology. Curr Oncol 2022;29(3):1828–39. https://doi.org/10.3390/curroncol29030150.

[11] Alexander M, Kim SY, Cheng H. Update 2020: management of non-small cell lung cancer. Lung 2020;198(6):897–907. https://doi.org/10.1007/s00408-020-00407-5.

[12] Lamb YN. Osimertinib: a review in previously untreated, EGFR mutation-positive, advanced NSCLC. Targ Oncol 2021;16(5):687–95. https://doi.org/10.1007/s11523-021-00839-w.

[13] Bulbul A, Husain H. FirsT-line treatment in EGFR mutant non-small cell lung cancer: is there a best option? Front Oncol 2018;8:94. https://doi.org/10.3389/fonc.2018.00094.

[14] Ahn M, Tsai C, Shepherd FA, et al. Osimertinib in patients with T790M mutation-positive, advanced non–small cell lung cancer: long-term follow-up from a pooled analysis of 2 phase 2 studies. Cancer 2019;125(6):892–901. https://doi.org/10.1002/cncr.31891.

[15] Du X, Shao Y, Qin HF, Tai YH, Gao HJ. ALK- rearrangement in non-small-cell lung cancer (NSCLC): ALK rearrangement in lung cancer. Thorac Cancer 2018;9(4): 423–30. https://doi.org/10.1111/1759-7714.12613.

[16] Arbour KC, Riely GJ. Diagnosis and treatment of anaplastic lymphoma kinase–positive non–small cell lung cancer. Hematol/Oncol Clin North Am 2017; 31(1):101–11. https://doi.org/10.1016/j.hoc.2016.08.012.

[17] Duruisseaux M, Besse B, Cadranel J, et al. Overall survival with crizotinib and next-generation ALK inhibitors in ALK-positive non-small-cell lung cancer (IFCT-1302 CLINALK): a French nationwide cohort retrospective study. Oncotarget 2017;8 (13). https://doi.org/10.18632/oncotarget.15746.

[18] Chaft JE, Rimner A, Weder W, Azzoli CG, Kris MG, Cascone T. Evolution of systemic therapy for stages I–III non-metastatic non-small-cell lung cancer. Nat Rev Clin Oncol 2021;18(9):547–57. https://doi.org/10.1038/s41571-021-00501-4.

[19] Reck M, Remon J, Hellmann MD. First-line immunotherapy for non–small-cell lung cancer. JCO 2022;40(6):586–97. https://doi.org/10.1200/JCO.21.01497.

[20] Clinical Trial Endpoints for the Approval of Non- Small Cell Lung Cancer Drugs and Biologics. Publ online 2015:19.

[21] Shameer K, Zhang Y, Jackson D, et al. Correlation between early endpoints and overall survival in non-small-cell lung cancer: a trial-level meta-analysis. Front Oncol 2021;11:672916. https://doi.org/10.3389/fonc.2021.672916.

[22] Bradburn MJ, Clark TG, Love SB, Altman DG. Survival Analysis Part II: multivariate data analysis – an introduction to concepts and methods. Br J Cancer 2003;89(3): 431–6. https://doi.org/10.1038/sj.bjc.6601119.

[23] Van Wijk RC, Simonsson USH. Finding the right hazard function for time- to-event modeling: a tutorial and shiny application. CPT Pharmacomet Syst Pharm 2022;11: 991–1001. https://doi.org/10.1002/psp4.12797.

[24] Royston P, Lambert PC. Flexible parametric survival analysis using stata: beyond the cox model. Stata Press,; 2011.

[25] Royston P, Parmar MKB. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. Stat Med 2002;21(15): 2175–97. https://doi.org/10.1002/sim.1203.

[26] Hofer F, Kauczor HU, Stargardt T. Cost-utility analysis of a potential lung cancer screening program for a high-risk population in Germany: a modelling approach. Lung Cancer 2018;124:189–98. https://doi.org/10.1016/j.lungcan.2018.07.036.

[27] Lai YH, Chen WN, Hsu TC, Lin C, Tsao Y, Wu S. Overall survival prediction of non-small cell lung cancer by integrating microarray and clinical data with deep learning. Sci Rep 2020;10(1):4679. https://doi.org/10.1038/s41598-020-61588-w.

[28] Zhao M, Pan X, Yin Y, et al. Cost-effectiveness analysis of five systemic treatments for unresectable hepatocellular carcinoma in china: an economic evaluation based on network meta-analysis. Front Public Health 2022;10:869960. https://doi.org/10.3389/fpubh.2022.869960.

[29] Planchard D, Popat S, Kerr K, et al. Metastatic non-small cell lung cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. Ann Oncol 2018;29:iv192–237. https://doi.org/10.1093/annonc/mdy275.

[30] Goldstraw P, Chansky K, Crowley J, et al. The IASLC lung cancer staging project: proposals for revision of the tnm stage groupings in the forthcoming (eighth) edition of the TNM classification for lung cancer. J Thorac Oncol 2016;11(1): 39–51. https://doi.org/10.1016/j.jtho.2015.09.009.

[31] Guyot P, Ades A, Ouwens MJ, Welton NJ. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. BMC Med Res Method 2012;12(1):9. https://doi.org/10.1186/1471-2288-12-9.

[32] Liu N, Zhou Y, Lee JJ. IPDfromKM: reconstruct individual patient data from published Kaplan-Meier survival curves. BMC Med Res Method 2021;21(1):111. https://doi.org/10.1186/s12874-021-01308-8.

[33] Rogula B, Lozano-Ortega G, Johnston KM. A method for reconstructing individual patient data from kaplan-meier survival curves that incorporate marked censoring times. MDM Policy Pract 2022;7(1). 238146832210776. doi:10.1177/23814683221077643.

[34] Wei Y., Royston P. Reconstructing time-to-event data from published Kaplan–Meier curves. Published online 2018:17.

[35] Hoyle MW, Henley W. Improved curve fits to summary survival data: application to economic evaluation of health technologies. BMC Med Res Method 2011;11(1): 139. https://doi.org/10.1186/1471-2288-11-139.

[36] Jackson C. flexsurv: a platform for parametric survival modeling in R. J Stat Softw 2016;70(8):1–33. https://doi.org/10.18637/jss.v070.i08.

[37] Kümmel A, Bonate PL, Dingemanse J, Krause A. Confidence and prediction intervals for pharmacometric models: confidence & prediction intervals for pharmacometrics. CPT Pharmacomet Syst Pharm 2018;7(6):360–73. https://doi.org/10.1002/psp4.12286.

[38] C.R.A.N. Documentation. rockchalk R package. CRAN.R. Published 2022. Accessed November 3, 2021. ⟨https://rdocumentation.org/packages/rockchalk/versions/1.8.157⟩.

[39] C.R.A.N. Documentation. rmst_generic R package. CRAN.R. Published 2022. Accessed November 3, 2021. ⟨https://www.rdocumentation.org/packages/flexsurv/versions/2.0/topics/rmst_generic⟩.

[40] Huwaldt J. PlotDigitizer. Published online 2018.

[41] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. Published 2020. Accessed May 15, 2021. ⟨https://www.R-project.org/⟩.

[42] Ou SHI, Ziogas A, Zell JA. AsiaN Ethnicity Is A Favorable Prognostic Factor for Overall Survival in Non-small Cell Lung Cancer (NSCLC) and Is Independent of Smoking Status. J Thorac Oncol 2009;4(9):1083–93. https://doi.org/10.1097/JTO.0b013e3181b27b15.

[43] Driessen E, Detillon D, Bootsma G, et al. Population-based patterns of treatment and survival for patients with stage I and II non-small cell lung cancer aged 65–74 years and 75 years or older. J Geriatr Oncol 2019;10(4):547–54. https://doi.org/10.1016/j.jgo.2019.03.001.

[44] Paz-Ares LG, Zimmermann A, Ciuleanu T, et al. Meta-analysis examining impact of age on overall survival with pemetrexed for the treatment of advanced non-squamous non-small cell lung cancer. Lung Cancer 2017;104:45–51. https://doi.org/10.1016/j.lungcan.2016.12.007.

[45] Kawaguchi T, Takada M, Kubo A, et al. Performance status and smoking status are independent favorable prognostic factors for survival in non-small cell lung cancer: a comprehensive analysis of 26,957 patients with NSCLC. J Thorac Oncol 2010;5 (5):620–30. https://doi.org/10.1097/JTO.0b013e3181d2dcd9.

[46] Mok T, Camidge DR, Gadgeel SM, et al. Updated overall survival and final progression-free survival data for patients with treatment-naive advanced ALK-positive non-small-cell lung cancer in the ALEX study. Ann Oncol 2020;31(8): 1056–64. https://doi.org/10.1016/j.annonc.2020.04.478.

[47] Ramalingam SS, Vansteenkiste J, Planchard D, et al. Overall survival with osimertinib in untreated, EGFR -mutated advanced NSCLC. N Engl J Med 2020;382 (1):41–50. https://doi.org/10.1056/NEJMoa1913662.
[48] 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): 1657–69. https://doi.org/10.1016/j.jtho.2020.06.015.
[49] Gadgeel S, Rodríguez-Abreu D, Speranza G, et al. Updated analysis from KEYNOTE-189: pembrolizumab or placebo plus pemetrexed and platinum for previously untreated metastatic nonsquamous non–small-cell lung cancer. JCO 2020;38(14):1505–17. https://doi.org/10.1200/JCO.19.03136.
[50] Reck M, Rodríguez-Abreu D, Robinson 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 2019;37(7):537–46.
[51] Dang TO, Ogunniyi A, Barbee MS, Drilon A. Pembrolizumab for the treatment of PD-L1 positive advanced or metastatic non-small cell lung cancer. Expert Rev Anticancer Ther 2016;16(1):13–20. https://doi.org/10.1586/14737140.2016.1123626.
[52] Kaprin A.D., Starinskii V.V., Shakhzadova A.O. The State of Oncology Assistance for the Population of Russia in 2021.; 2022:239. ⟨https://oncology-association.ru/wp-content/uploads/2022/05/sostoyanie-onkologicheskoj-pomoshhi-naseleniyu-rossii-v-2021-godu.pdf⟩.
[53] 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. https://doi.org/10.1001/jamaoncol.2021.4932.
[54] Van Meerbeeck JP, Franck C. Lung cancer screening in Europe: where are we in 2021? Transl Lung Cancer Res 2021;10(5):2407–17. https://doi.org/10.21037/tlcr-20-890.
[55] del Campo C, Bai J, Keller LR. Comparing markov and non-markov alternatives for cost-effectiveness analysis: insights from a cervical cancer case. Oper Res Health Care 2019;21:32–43. https://doi.org/10.1016/j.orhc.2019.04.001.
[56] Filipović-Pierucci A., Zarca K., Durand-Zaleski I. Markov Models for Health Economic Evaluations: The R Package heemod. arXiv:170203252 [stat]. Published online April 25, 2017. Accessed February 20, 2020. ⟨http://arxiv.org/abs/1702.03252⟩.
[57] Cambridge WA, Fairfield C, Powell J, et al. Meta-analysis and meta-regression of survival after liver transplantation for unresectable perihilar cholangiocarcinoma. Ann Surg 2021;237(2):240–50. https://doi.org/10.1097/SLA.0000000000003801.
[58] Li M, Dave N, Salem AH, Freise KJ. Model-based meta-analysis of progression-free survival in non-Hodgkin lymphoma patients. Medicine 2017;96(35):e7988. https://doi.org/10.1097/MD.0000000000007988.
[59] Tran KA, Kondrashova O, Bradley A, Williams ED, Pearson JV, Waddell N. Deep learning in cancer diagnosis, prognosis and treatment selection. Genome Med 2021;13(1):152. https://doi.org/10.1186/s13073-021-00968-x.
[60] Qiao N, Insinga R, De Lima Lopes Junior G, Cook J, Sénécal M. A review of cost-effectiveness studies of pembrolizumab regimens for the treatment of advanced non-small cell lung cancer. Pharm Open 2021;5(3):365–83. https://doi.org/10.1007/s41669-020-00255-2.
[61] Chouaïd C, Molinier L, Combescure C, Daurès JP, Housset B, Vergnenègre A. Economics of the clinical management of lung cancer in France: an analysis using a Markov model. Br J Cancer 2004;90(2):397–402. https://doi.org/10.1038/sj.bjc.6601547.
[62] Hirst TC, Sena ES, Macleod MR. Using median survival in meta-analysis of experimental time-to-event data. Syst Rev 2021;10(1):292. https://doi.org/10.1186/s13643-021-01824-0.
[63] Dearden S, Stevens J, Wu Y-L, et al. Mutation incidence and coincidence in non small-cell lung cancer: meta-analyses by ethnicity and histology (mutMap). Ann Oncol 2013;24(9):2371–6. https://doi.org/10.1093/annonc/mdt205.
[64] Zhudenkov K., Katuninks N., Shulgin B., Peskov K. A model-based survival meta-analysis for indirect comparison of immune therapy efficacy in NSCLC. Poster presented at: Annual Meeting of the Population Approach Group in Europe; June 26–28, 2024; Rome, Italy.
[65] Turner D.C., Wada R., Zhou H., et al. Model-based meta-analysis of non-small cell lung cancer with standard of care PD -1 inhibitors and chemotherapy for early development decision making. CPT Pharmacom & Syst Pharma. Published online January 31, 2023:psp4.12917. doi:10.1002/psp4.12917.
[66] Duma N, Santana-Davila R, Molina JR. Non–small cell lung cancer: epidemiology, screening, diagnosis, and treatment. Mayo Clin Proc 2019;94(8):1623–40. https://doi.org/10.1016/j.mayocp.2019.01.013.
[67] Jabbour SK, Lee KH, Frost N, et al. Pembrolizumab plus concurrent chemoradiation therapy in patients with unresectable, locally advanced, stage III non–small cell lung cancer: the phase 2 KEYNOTE-799 nonrandomized trial. JAMA Oncol 2021;7 (9):1351. https://doi.org/10.1001/jamaoncol.2021.2301.
[68] Shen D, Wang J, Wu J, et al. Neoadjuvant pembrolizumab with chemotherapy for the treatment of stage IIB–IIIB resectable lung squamous cell carcinoma. J Thorac Dis 2021;13(3):1760–8. https://doi.org/10.21037/jtd-21-103.
[69] Spigel DR, Faivre-Finn C, Gray JE, et al. Five-Year Surviv Outcomes Pac Trial: Durvalumab Chemoradiotherapy Stage III Non-Small-Cell Lung Cancer [Publ Correct Appears J Clin Oncol 2022 Jun 10;40(17):1965. doi: 10.1200/JCO.22.01023]. J Clin Oncol. 2022;40(12):1301-1311. doi:10.1200/JCO.21.01308.
[70] Filippi AR, Bar J, Chouaid C, 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;9(6):103464. https://doi.org/10.1016/j.esmoop.2024.103464.
[71] Iyengar P, Zhang-Velten E, Court L, et al. Accelerated hypofractionated image-guided vs conventional radiotherapy for patients with stage II/III non–small cell lung cancer and poor performance status: a randomized clinical trial. JAMA Oncol 2021;7(10):1497. https://doi.org/10.1001/jamaoncol.2021.3186.
[72] Vinod SK, Hau E. Radiotherapy treatment for lung cancer: current status and future directions. Respirology 2020;25(S2):61–71. https://doi.org/10.1111/resp.13870.
[73] Paracha N, Abdulla A, MacGilchrist KS. Systematic review of health state utility values in metastatic non-small cell lung cancer with a focus on previously treated patients. Published 2018 Sep 12. Health Qual Life Outcomes 2018;16(1):179. https://doi.org/10.1186/s12955-018-0994-8.
[74] Chen W, Li L, Ji S, Song X, Lu W, Zhou T. Longitudinal model–based meta-analysis for survival probabilities in patients with castration-resistant prostate cancer. Eur J Clin Pharm 2020;76(4):589–601. https://doi.org/10.1007/s00228-020-02829-9.