A Description of the IVI-NSCLC Model v1.0

Devin Incerti*   Jeroen P. Jansen*

December 12, 2018

Contents

*Innovation and Value Initiative

List of Figures

List of Tables

1   Open Source Value Project

The continuing increase in U.S. healthcare costs has stimulated the introduction of initiatives to promote the use of high-value care. Cost-effectiveness analysis can inform efficient use of healthcare resources by formally computing costs and benefits to identify the most valuable treatment options for a given disease. In many countries, a single health technology assessment agency assesses the value of healthcare technology by means of cost-effectiveness analysis and recommends a utilization strategy. In the US, however, utilization decisions are decentralized and made by a variety of payers and provider organizations. Value frameworks are gaining prominence to guide utilization of therapies, but vary in perspective, the evidence considered, and approaches, thereby resulting in confusion and debate among stakeholders.

A thorough evidence-based analysis of the value of medical technology is resource intensive and complex. Typically, there is no empirical study with sufficient long-term follow-up that compares all treatments for a particular disease regarding relevant clinical outcomes and costs. Thus, cost-effectiveness analyses generally rely on mathematical models that integrate evidence on the course of disease, treatment effects, and the relationship between clinical outcomes and costs, from a variety of studies. The nature of these evaluations can lead to disputes in the scientific literature and community. Models are typically difficult to understand. Even modeling experts may not be able to fully understand a model-based cost-effectiveness analysis without public source code and detailed model documentation. This lack of transparency also poses problems for users whose perspective, local context, or patient population varies from that of the reported analysis. In the absence of public access to the actual model, updating the evaluation is cumbersome, if not impossible, for someone other than the original model developer. As a result, published cost-effectiveness findings risk immediate irrelevance to some stakeholders and growing irrelevance to all stakeholders as new clinical evidence emerges. Value assessment only has relevance for decision-making when it reflects the totality of the latest evidence, is transparent, deemed credible by different stakeholders, representative of the local context and patient population, and can be easily updated without duplication of effort.

With the Open-Source Value Project (OSVP), the Innovation and Value Initiative (IVI) aims to maximize both the relevance and credibility of value assessment in the context of the United Statesâ decentralized decision-making environment by developing and providing access to flexible open-source decision models for value assessment. These interactive models have two primary objectives: (i) to enable a more constructive dialogue regarding value assessment between stakeholders (e.g. patients, payers, providers, and manufacturers) with different beliefs about relevant clinical data, modeling approaches, and value perspectives; and (ii) to provide local decision-makers with means to credible value assessment that reflects the local setting and is based on the latest evidence while accounting for all scientific uncertainty (due to patient heterogeneity, gaps in evidence, and different modeling beliefs).

In order for a decision model to remain relevant over time it needs to evolve along with the clinical evidence and scientific insights. OSVP facilitates iterative development and collaboration between multiple clinical and methodological experts. Refer to the IVI website here for additional information about the OSVP.

2   Topic definition

IVI was interested in developing its most recent OSVP within oncology. When selecting the tumor type of interest, IVI considered criteria such as burden of disease, development of innovative treat-

ment, alternative treatment strategies available, availability of clinical evidence, and engagement of patient organization(s) to actively contribute to the project.

Lung cancer is the leading cause of cancer related death worldwide (**?**). Non-small lung cancer (NSCLC) accounts for an estimated 85% of lung cancer cases and comprises adenocarcinoma, squamous cell carcinoma, and large cell carcinoma (**?**). The five-year survival of stage IV NSCLC is less than 2% (**?**). Given the rapid pace of development in NSCLC, the scope of the OSVP needed to be limited to a very specific sub-population in order to ensure a model could be developed in a reasonable time frame that allowed IVI to demonstrate the typical areas of uncertainty in value assessment in oncology. The selected target population of interest for the most recent OSVP is metastatic epidermal growth factor receptor positive (EGFR+) NSCLC. EGFR mutations are more commonly observed in tumors from female patients with adenocarcinomas without a history of smoking and with Asian ethnicity, but can occur in patients with prior smoking history and across all races and genders (**?**). The evidence base for the treatments used for the EGFR+ population is more modest than the evidence base for treatments used for EGFR negative NSCLC, which makes development of a model reflective of the latest evidence base more manageable given the logistical constraints set for this project. Future activities may include expanding the model to other subpopulations of interest.

3   Purpose

The aim was to develop a flexible open-source simulation model that can be used to estimate the value of alternative sequential treatment strategies for patients with metastatic EGFR+ NSCLC. The IVI-NSCLC(egfr+) model is accessible to both technical and non-technical end-users and allows them evaluating the impact of uncertainty in clinical evidence, alternative model structures, the decision framework of choice (i.e. cost-effectiveness analysis [CEA]) or multi-criteria decision analysis [MCDA]), novel concepts of value, and perspective (healthcare or limited societal) on the estimates of value.

3.1   Value assessment

The IVI-NSCLC(egfr+) model is designed to assess the value of multiple competing sequential treatment strategies for patients with metastatic EGFR+ NSCLC starting with 1st line treatment (1L), followed by 2nd line treatment (2L), and treatment beyond 2nd line (2L+) as outlined in Figure 1. Sequences that can be evaluated with version 1 of the model were informed by the available evidence base, guidelines, and clinician input.

The IVI-NSCLC(egfr+) model is suitable for informing decisions for specific (sub)populations, but is not suitable for making predictions at the individual level, nor should it replace the patient-physician shared decision-making process. Local decision makers can modify the model to perform analysis of value that reflect the local setting while accounting for all scientific uncertainty and help them understand the confidence with which they make decisions.

The IVI-NSCLC(egfr+) model is not a value assessment framework, but a model that simulates the costs, health outcomes, and risks associated with sequential treatment sequences for metastatic EGFR+ NSCLC. It can therefore be used with any value framework preferred by the user. Currently, two methodologies for decision analysis are supported by the model: CEA based on cost per quality adjusted life year (QALY) expressed as net-monetary benefit (NMB) and MCDA (**?**).

The MCDA was implemented according to **?** and based on the following criteria: progression free survival (PFS), overall survival (OS), adverse events, expected time on first, second, and third line
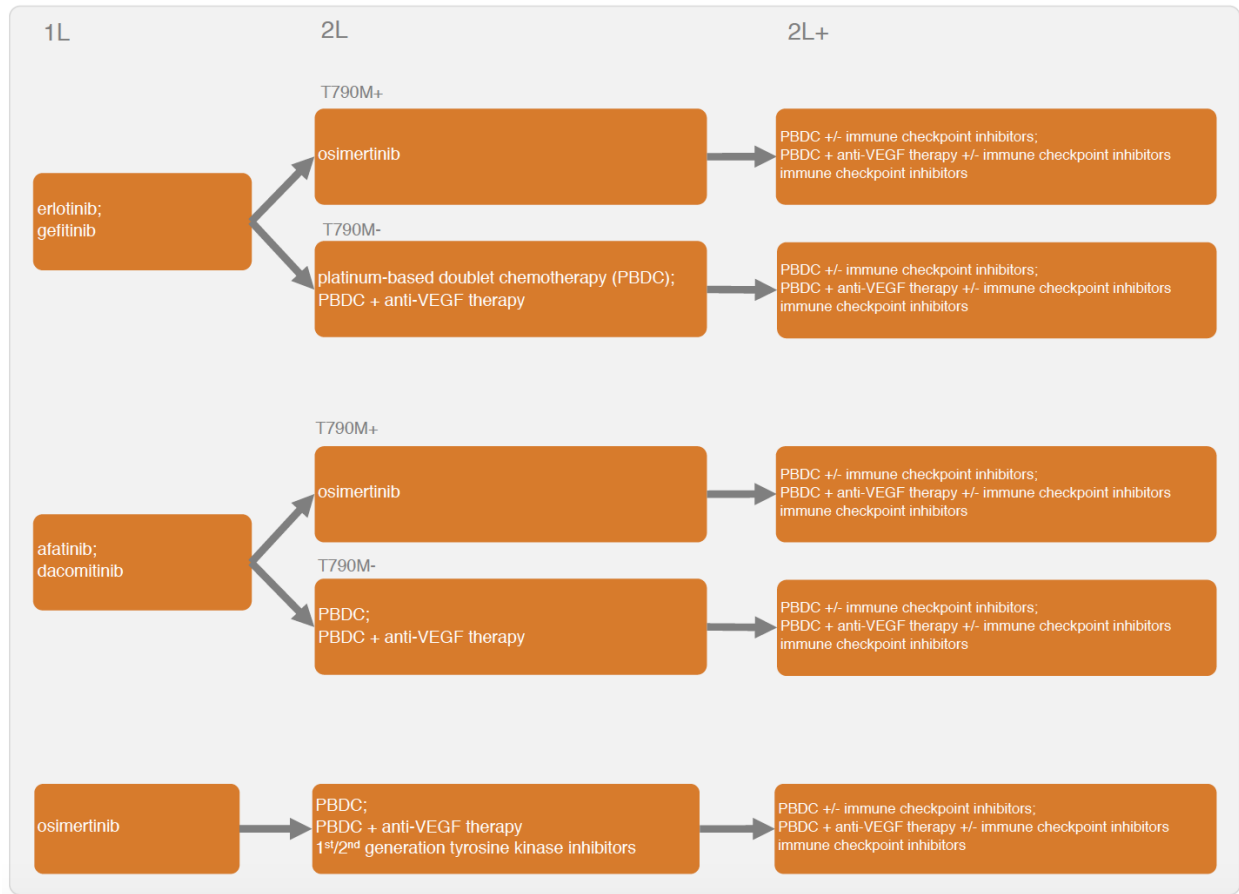
Figure 1: Sequential treatment strategies of interest to be compared with the model

treatment, utility/quality of life, health care sector costs, productivity losses, route of administration (oral/injection/infusion) and time the medication has been on the market. The assessment of value can be performed from a health care sector perspective by only incorporating health care sector costs, or from a (limited) societal perspective by including productivity losses in addition.

? suggest five concepts of value that researchers should consider adding to the standard cost per QALY based CEA: (i) a reduction in uncertainty from a diagnostic test; (ii) insurance value for healthy individuals due to reduction against physical risk; (iii) the value of hope for individuals who become risk-loving and would rather pay for a therapy with a long right survival tail than a therapy with a shorter right survival tail but an equivalent (or shorter) expected life-expectancy; (iv) real option value when a therapy allows an individual to benefit from future medical innovations; and (v) scientific spillovers when the benefits of an innovation cannot be entirely appropriated by the innovator.

The value of hope is most relevant for innovations that increase longevity and might be particularly well suited to the analyses of treatments for NSCLC. Traditionally, CEA focuses on maximizing expected QALYs; however, patients might be willing to take gambles (i.e., they become "risk lovers") and care about the variation in benefits and costs, not simply the means. If patients value hope, they may prefer the treatment with greater variability in survival over the treatment with less variability

despite having the same expected survival. In contrast, if patients are risk-averse, they may prefer the latter to avoid an unlucky outcome. Either way, they have a preference for one intervention over its alternative that appear identical based on its average costs and benefits. Patients may place substantial value on a modest chance of a durable survival response, over and above average survival, and decision-makers acting on their behalf may want to consider this aspect when making population level decisions regarding the value of interventions.

The IVI-NSCLC(egfr+) model allows users to incorporate value of hope into their analyses, while noting that the approach is less well-established than conventional cost-effectiveness analysis. Future versions of the IVI-NSCLC(egfr+) may consider incorporating other value components, such as insurance value.

3.2 Evaluation of scientific uncertainty

Decision models can be used to inform efficient use of health care resources, but often lead to scientific disagreements and mistrust among stakeholders. Model inputs are typically informed by a formal evidence synthesis to ensure all relevant evidence is considered. However, decisions regarding the mathematical structure relating model inputs to outputs are frequently made arbitrarily based on the idiosyncratic expertise of the model developer without evaluating the impact on findings. While robustness can be assessed by means of sensitivity analyses, these are typically limited to studying the impacts of varying model inputs. For any given disease, a variety of modeling approaches have typically been proposed in the literature. In order to evaluate the impact of these different approaches on estimates of value in a systematic way, we need flexible open-source models to not only capture the uncertainty in model input parameters (i.e. parameter uncertainty), but also capture alternative model structures (i.e. structural uncertainty). This flexibility facilitates demonstrating the implications of different areas of uncertainty and leads to a better understanding of the reasons why value estimates can vary.

4 Iterative process

In order for a decision model to remain relevant over time it needs to evolve along with the clinical evidence and scientific insights. An open-source approach facilitates iterative development and collaboration between multiple clinical and methodological experts. We will use a four-step process for the development of flexible decision or simulation models for value assessment:

1. Public release of the model. The initial release of an OSVP model must be flexible and allow users to choose from a large number of plausible model structures and approaches based on clinical practice and previous modeling efforts.

2. Invite feedback and suggested improvements to the model in a public comment period.

3. A panel of experts determines which of the evidence-based suggestions for improvement suggested in Step 2 should be implemented by means of a modified Delphi process.

4. Revise the model based on the feedback from the technical expert panel in Step 3. The four-step process is designed to be repeated many times so that the modeling approach and evidence considered can be refined over time.

5 Components

Version 1 of the IVI-NSCLC(egfr+) model is designed to provide a starting point for open debate. To facilitate transparency, understanding, debate and collaboration among diverse stakeholders, the

IVI-lung model will consist of the following components available in the public domain:

Source code: R and C++ code for the model are available in an IVI GitHub repository. Modelers and programmers may adapt the source code for their own purposes or collaborate to improve the code.

R package: The model will be released as an R-package with documentation available online. Researchers can use the package to run the model for custom analyses. Use of the package is recommended when performing analyses for scientific research.

Detailed model documentation: This document provides extensive technical details on the model structure, statistical methods for parameter estimation, and source data.

Expert Model Interface: For users not well-versed in the programming language, a web application for running the model online is provided. The web application is designed for custom analyses and allows users full control over the treatments, patient population, model structures, parameter values, and simulation settings.

Value Tool: An important aim of OSVP is to obtain feedback from as many relevant stakeholders as possible. A general audience web-application has been developed, allowing those who are not experts in modeling or health economics, to interact with the model.

6 Model structure

6.1 Disease model

The IVI-NSCLC is an individual-level continuous-time state transition model (CTSTM) in which patients can either have stable disease, progressed disease, or have died. To model sequential treatment strategies, we assumed that patients move to the next treatment after disease progression. In other words, a patient with stable disease moves to the second treatment in a sequence after progression on the first treatment, to the third treatment after progression on the second treatment, and so on. A patient can die at any time. Figure 2 shows health state transitions for an example sequence of 3 treatments. Sequential treatment is incorporated into the CTSTM by expanding the number of health states according to the number of treatment lines. In the figure there are 3 treatment lines and 5 health states; more generally, there is one health state for each treatment line, a health state after progression on the final line, and a death state, so a model with n treatment lines will have n+2 health states.

The hazard rate, h^{qr} (u) for a transition from state q to state r at time u follows different parametric distributions including the exponential, Weibull, Gompertz, log-logistic, and fractional polynomial distributions as estimated with a multi-state network meta-analysis (NMA). In the individual-level model, time to each state r that can be entered from state q is sampled based on time-varying hazard rates and patients' transition to the state with the shortest sampled time.

The parameterization of transitions between health states allows incorporating potentially relevant prognostic factors or effect-modifiers. This would also facilitate including carry-over effects from one line of treatment to the next, once such data becomes available.
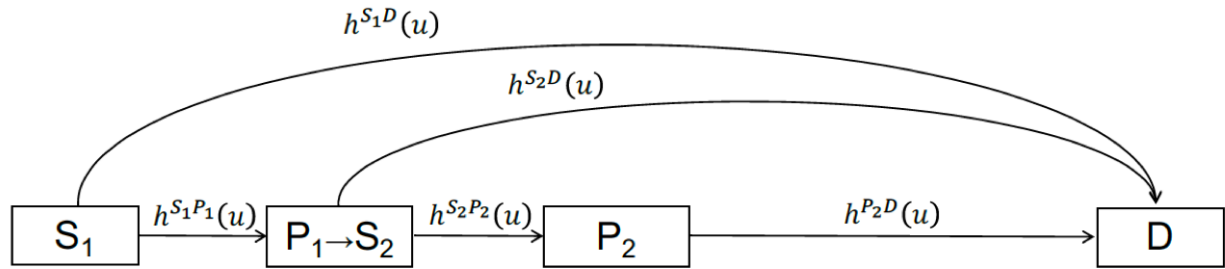
S_1 = Progression-free (stable disease) with 1L treatment

P_1 = Progression with 1L treatment

S_2 = Progression-free (stable disease) with 2L treatment

P_2 = Progression with 2L treatment, captures the survival with 2L+ without making a distinction between a progression free and progression phase

D = Dead

$h^{S_1 P_1}(u)$ = hazard for transitioning from progression-free to progression with 1L treatment at time u

$h^{S_1 D}(u)$ = hazard for transitioning from progression-free to dead with 1L treatment at time u

$h^{S_2 P_2}(u)$ = hazard for transitioning from progression-free to progression with 2L treatment at time u

$h^{S_2 D}(u)$ = hazard for transitioning from progression-free to dead with 2L treatment at time u

$h^{P_2 D}(u)$ = hazard for transitioning from progression on 2L to dead at time u

Figure 2: Model structure with 4 states describing development of disease over time for a sequence starting with 1L, followed by 2L and 2L+ treatment; 2L+ treatment is captured with the L2 progression state

S_1 = Progression-free (stable disease) with 1L treatment

P_1 = Progression with 1L treatment, captures the survival with 2L and 2L+ without making a distinction between progression free and progression phases

D = Dead

$h^{S_1P_1}(u)$ = hazard for transitioning from progression-free to progression with 1L treatment at time u
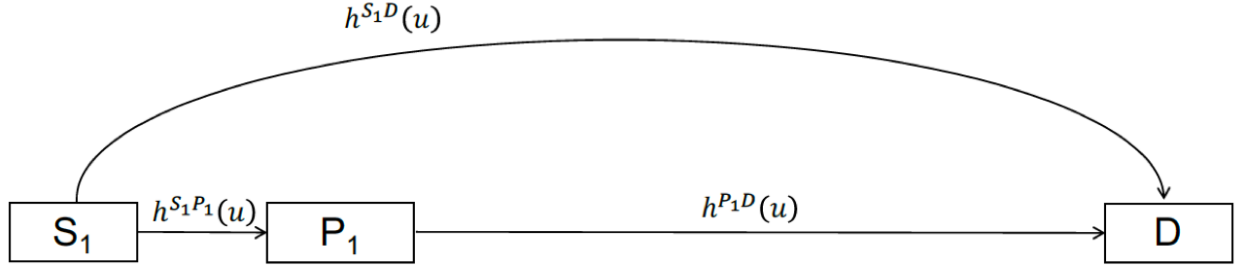
$h^{S_1D}(u)$ = hazard for transitioning from progression-free to dead with 1L treatment at time u

$h^{P_1D}(u)$ = hazard for transitioning from progression on 1L to dead at time u

Figure 3: Alternative model structure with 3 states describing development of disease over time for a treatment sequence starting with 1L until death; 2L and 2L+ treatment is captured with the progression state

6.2 Adverse events

The model includes multiple distinct adverse events associated with oncology treatment by treatment line. More precisely, for each treatment line and each patient, the model samples whether each adverse event occurs from a binomial distribution.

6.3 Cost and utility

All non-death health states are associated with utility and cost values. Cost values are separated into distinct categories (e.g., drug acquisition and administration costs, costs due to inpatient hospitalizations, etc.). Because we did not find strong evidence to the contrary, we assumed that utility and cost values remain constant over time within a given health state, but the model is also flexible enough to estimate utility and cost values that vary over time in a general manner. Adverse events cause utility decrements and more serious adverse events–such as those that require hospitalizations–increase health care costs.

Drug acquisition and administration related costs are modeled based on standard clinical practice. For example, consider a case in which a therapy is dosed daily. Then costs will accrue each day until disease progression and the patient switches to the next treatment in the sequence. Another possible scenario is that dosing is based on fixed cycles. In these cases, costs will accrue until the end of a cycle or until the patient switches treatment due to disease progression. Other scenarios will be considered as required.

The value of hope is illustrated with the model assuming representative relationships between utility values for each of the health states as a function of time reflecting risk-seeking and risk-averse behavior.

Adverse events cause utility decrements and more serious adverse events–such as those that require hospitalizations–increase health care costs.

6.4 Patient heterogeneity

The model is sufficiently flexible so that all model parameters (i.e., the parameters of probability distributions characterizing health state transitions, adverse event rates, costs, and utility) can depend on patient characteristics. In other words, we modeled input parameters as a function of covariates so that they can vary across individuals and treatments. That is, a given parameter, a_{jt}, for individual j using treatment t is modeled as $\alpha_{jt} = g^{-1}(X_{jt}\beta)$ where g^{-1} (.) is an inverse transformation function. However, the extent heterogeneity can be modeled depends on the level of detail in the available evidence. We modeled heterogeneity based off of the level of evidence available.

6.5 Rationale for individual-level simulation

Our primary motivation for using an individual-level model relates to the distinction between "clock forward" and "clock reset" approaches to parameter estimation. In the "clock reset" approach, time u in h^{qr} (u) resets to 0 after each transition whereas in the "clock forward" approach, time u refers to time since the start of the model. If a "clock forward" approach is used, then health state probabilities can be calculated analytically using the **?** estimator and a cohort CTSTM can be used; conversely, if a "clock reset" approach is taken, then individual-level simulation-based approaches must be used to estimate health state probabilities (**?**), (**?**), (**?**). In our case, each line specific multi-state NMA is a "clock-forward" model; however, when modeling sequential treatment multiple multi-state NMAs are used so the disease model is a mixture between the two: a "clock forward" model is used for each treatment but time u resets each time a patient begins a new treatment. To accommodate a mixture of "clock reset" and "clock forward" approaches, we use an individual-level model.

7 Model outcomes

The model simulates the health outcomes, risks, and costs associated with alternative treatment strategies in oncology. The model's time horizon can be selected by the user and will default to a lifetime horizon. The health outcomes, risks, and costs will be combined to assess value using two frameworks: CEA and MCDA. Both disaggregated model outcomes and value estimates will be based on mean values (i.e., averages across all simulated patients). An overview of all model outcomes is shown in Table 1. Additional details are provided in the text below.

7.1 Health outcomes

The primary health outcome is the probability that a patient is in a given health state at a given point in time following treatment initiation. A survival curve was generated from the health state probabilities as one minus the probability that a patient has died. Since the model is an individual patient simulation, life expectancy is just the average time of death across simulated patients. Life-years was calculated as age at death minus age at treatment initiation. QALYs are weighted life-years, with weights equal to the utility values assigned to states. Discounted QALYs were also calculated using a default discount rate of 3 percent.

Table 1: Model outcomes

Category	Outcomes
Health outcomes	Health state probabilities; progression free survival & overall survival; quality-adjusted life-years
Risks	Adverse events (diarrhea, dry skin, elevated alanine transaminase, elevated aspartate transaminase, eye problems, paronychia, pneumonitis, pruritus, rash, stomatitis)
Health care sector costs	Drug acquisition and administration costs; adverse event costs; costs from inpatient hospital stays; costs from hospital outpatient or doctor office visits
Non-health care sector costs	Productivity losses
Value assessment	Cost-effectiveness analysis; multi-criteria decision analysis

7.2 Risks

The model calculates the expected number of adverse events per patient as the mean number of adverse events experienced by patients during the simulation. To improve interpretability, this number is reported as the expected number of adverse events per 1,000 patients. Both the expected number of total adverse events and the expected number of adverse events by type are reported.

7.3 Costs

Costs are calculated for multiple categories and separated into health care sector and non-heath care sector costs as recommended by the ?. Health care sector cost categories included drug acquisition and administration costs, adverse events costs, the costs of inpatient hospital stays, and costs from hospital outpatient and doctor office visits. Non-health care sector costs include productivity losses. We used both human-capital approach and friction-cost approach to estimate productivity losses associated with the different treatment sequences. The level of detail regarding productivity related cost estimates that were incorporated is based on the available data in the literature.

7.4 Value assessment

Value can be assessed using two frameworks: CEA and MCDA. If CEA is used for value assessment, then the value of treatment is estimated using the NMB. CEA from a societal perspective includes productivity losses while analyses from a health care sector perspective do not.

With MCDA, the value of each treatment strategy is based on a "total value" score. The score is calculated in three steps. First, a number of distinct criteria are assigned values on a common scale, for instance, ranging from 0 to 100. Second, each criterion is assigned points, say, ranging from 0 to 10, which is, in turn, used to calculate a weight by dividing each criterion's points by the sum of points across all criteria. For example, if there were 3 criteria and each criterion was given a score of 5, then each criterion would receive a weight of 1/3. If, on the other hand, the three criteria were

given scores of 2.5, 5, and 7.5, then they would be given weights of 0.167, 0.33, and 0.5, respectively. Third, to aggregate results, we multiplied the value of each criterion on the common scale by its weight and sum the weighted scores across criteria. Future iterations of the model will consider using the Patient Experience Study to inform the baseline weights across attributes.

8 Source data and parameter estimation

Key parameters for the model relate to: (i) treatment effects in terms of transitions between the health states; (ii) adverse events; (iii) utilities; (iv) healthcare resource use; and (v) productivity. Parameter estimates for the current version of the IVI-NSCLC(egfr+) model are based on currently available published evidence identified by means of a systematic literature review and synthesized with meta-analysis techniques where appropriate.

8.1 Treatment effects for transition rates

Relevant evidence to estimate treatment effects in terms of transitions between the health states was obtained with a systematic literature review of randomized controlled trials (RCTs) evaluating the efficacy of relevant competing interventions for the treatment of adult patients with metastatic EGFR+ non-squamous NSCLC by line of treatment (i.e. 1L, 2L, and 2L+) in terms of OS and PFS. Details regarding eligibility criteria that defined the scope of the studies considered relevant are outlined in the Appendix. The identified evidence base was used to 1) estimate the relative treatment effects for each 1L treatment versus a defined reference treatment; 2) estimate the absolute effects with the corresponding 1L reference treatment; and 3) estimate the absolute treatment effects with 2L treatment options.

8.1.1 Relative treatment effects with 1L therapy

Relative treatment effects between the competing interventions for the 1L treatment of EGFR+ NSCLC were estimated by means of a Bayesian network meta-analysis (NMA) using alternative fixed and random effects models.

RCTs are considered the gold standard to assess treatment effects of medical interventions. For many disease states there are multiple competing interventions to choose from. However, an RCT will rarely include all competing interventions of interest for a particular disease state when there are more than two relevant treatment options available. As such, an individual trial rarely provides all the information needed to guide evidence-based treatment selection. When each RCT compares only a subset of the interventions of interest â it may be possible to represent the evidence base as a network where all trials have at least one intervention in common with another trial. Such a network of trials involving treatments compared directly, indirectly, or both, can be synthesized by means of a network meta-analysis (add reference). As a result we not only get pooled results of available treatment comparisons studied in a head-to-head fashion but also relative treatment effects between interventions not compared directly. In order to ensure that a network meta-analysis are not affected by bias due to differences in prognostic factors between studies, we want to only consider the relative treatment effects of each trial. This implies that all interventions have to be part of one network of trials where each trial has at least one intervention in common with another trial.

Patients are randomized only within trials, not across trials, so there is a risk that patients participating in different trials differ with respect to demographic, disease or other characteristics. In addition, features of the trials themselves may differ. If these trial or patient characteristics are

effect modifiers, i.e. they affect the treatment effects of an intervention versus a control, then there are systematic differences in treatment effects across trials. Systematic differences in known and unknown effect-modifiers among studies comparing the same interventions in direct fashion result in between-study heterogeneity. An imbalance in the distribution of effect modifiers between studies comparing different interventions will result in transitivity or consistency violations and therefore biased indirect comparisons.

The RCTs that were part of one evidence network and were deemed sufficiently similar were included in the NMA. The evidence network for 1L treatments is presented in (Appendix X), along with information on the distribution of patients characteristics across the studies in the network.

8.1.1.1 Network meta-analysis model

Traditionally, NMAs in oncology are based on separate analyses for OS and PFS, with relative treatment effects estimated using reported hazard ratios (HR). There are two limitations of this approach. First, HR estimates rely on the proportional hazard assumption, which is implausible if the hazard functions of the competing interventions cross (**?**). Second, separate analyses of OS and PFS ignore the structural relationship between stable disease, progression, and death and implicitly assume that transition probabilities for stable to death and progression to death are equal.

We overcome these limitations by introducing a multi-state NMA that explicitly estimates each possible transition in a 3-state model—stable to progression, stable to death, and progression to death–and modeling time-varying hazard rates and relative treatment effects with known parametric survival functions or fractional polynomials. Our approach is illustrated in Figure 4 and expressed mathematically in Equation 1.

$$
\ln\left(h_{ik}^{SP}(u)\right) = \begin{cases} \alpha_{1ik} + \alpha_{2ik}u^{p_1} + \alpha_{3ik}u^{p_2} & \text{if } p_1 \neq p_2 \\ \alpha_{1ik} + \alpha_{2ik}u^{p} + \alpha_{3ik}u^{p}\ln(u) & \text{if } p_1 = p_2 = p \end{cases}
$$
$$
\ln\left(h_{ik}^{SD}(u)\right) = \alpha_{4ik}
$$
$$
\ln\left(h_{ik}^{PD}(u)\right) = \alpha_{5ik} + \alpha_{6ik}u^{p_1}
$$

$$
\begin{pmatrix} \alpha_{1ik} \\ \alpha_{2ik} \\ \alpha_{3ik} \\ \alpha_{4ik} \\ \alpha_{5ik} \\ \alpha_{6ik} \end{pmatrix} = \begin{pmatrix} \mu_{1ik} \\ \mu_{2ik} \\ \mu_{3ik} \\ \mu_{4ik} \\ \mu_{5ik} \\ \mu_{6ik} \end{pmatrix} + \begin{pmatrix} \delta_{1,ik} \\ d_{2,t_{ik}} - d_{2,t_{i1}} \\ 0 \\ 0 \\ d_{3,t_{ik}} - d_{3,t_{i1}} \\ 0 \end{pmatrix} \tag{1}
$$

$$
\delta_{1,ik} \sim N(d_{1,t_{ik}} - d_{1,t_{i1}}, \sigma_{d_1}^2)
$$

where $u^0 = \ln(u)$ and $d_{1,1} = 0$, $d_{2,1} = 0$, and $d_{3,1} = 0$

$h_{ik}^{SD}(u)$, $h_{ik}^{PD}(u)$, and $h_{ik}^{SD}(u)$ are the the hazard rate for disease progression, dying post-progression, and dying pre-progression, respectively, in study i for treatment arm k at time u. The $\alpha_{\cdot ik}$ are
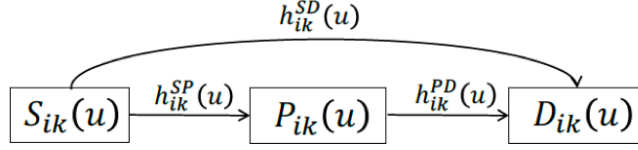
$S_{ik}(u)$ = progression -free (stable disease) in study i, treatment arm k at time u

$P_{ik}(u)$ = progressed disease in study i, treatment arm k at time u

$D_{ik}(u)$ = dead in study i, in treatment arm k at time u

$h_{ik}^{SP}(u)$ = hazard rate for disease progression in study i, in treatment arm k at time u

$h_{ik}^{PD}(u)$ = hazard rate for dying post-progression in study i, in treatment arm k at time u

$h_{ik}^{SD}(u)$ = hazard rate for dying pre-progression in study i, in treatment arm k at time u

Figure 4: Relationship between stable disease (S), progression (P) and death (D) as used in the multi-state network meta-analysis model

regression coefficients that represent the scale and shape parameters of the log hazard function in study i for treatment arm k. The $\mu_{\cdot i}$ reflect the study effects regarding the scale and shape parameters in each study i. The $\delta_{1,ik}$ are the study specific true underlying relative treatment effects regarding the scale of the log hazard function for the transition from stable to progression, which are described by a normal distribution with an average effect for treatment t relative to reference treatment 1, $d_{1,1t}$, and between study heterogeneity $\sigma_{d_1}^2$. $d_{2,1t}$ are the treatment specific relative effects regarding the first shape parameter of the log hazard function for the transition from stable to progression relative to reference treatment 1. To facilitate parameter estimation, we assumed that treatment has no effect on the 2nd shape parameter. Transition rates between stable disease and death are likely to be relatively small and therefore assumed to be independent of time and the same for all treatments. $d_{3,1t}$ are the treatment specific relative effects regarding the scale parameter of the log-hazard function for the transition from progression to death relative to reference treatment 1.

When $p_1 = 0$ and $\alpha_{3ik} = 0$ the log-hazard functions for the transitions between stable disease and progression, and progression and death follow a Weibull distribution. When $p_1 = 1$ and $\alpha_{3ik} = 0$ these log-hazard functions follow a Gompertz distribution. When $(p_1, p_2) = (0,0), (0,1)\}$ and $\alpha_{3ik} \neq 0$, the log-hazard functions follow a second order fractional polynomial that are extensions of the Weibull and Gompertz model to for allow arc- and bathtub shaped log-hazard functions.

A fixed effects model is obtained by replacing $\delta_{1,ik} \sim N(d_{1,t_{ik}} - d_{1,t_{i1}}, \sigma^2)$ with $\delta_{1,ik} = d_{1,t_{ik}} - d_{1,t_{i1}}$.

If it is assumed that treatment has only a direct effect on the transitions from stable to progression, which can be reasonably defended when a particular treatment upon disease progression is discontinued, the model can be simplified by setting $d_{3,t_{i1}} = 0$.

8.1.1.2   Likelihood

The model parameters were estimated based on the number of patients r in each of the three health states (stable, progression and death) at time u for each arm k of each trial i obtained from the published Kaplan-Meier (KM) curves for time points for which both PFS and OS were reported.

Accordingly, a multinomial likelihood was used for the proportion of patients in each of the three health states at any time point for each arm of each trial ($S_{ik}(u)$, $P_{ik}(u)$, and $D_{ik}(u)$) according to Equation 2.

$$(r_{ik}^S(u), r_{ik}^P(u), r_{ik}^D(u)) \sim multinomial(S_{ik}(u), P_{ik}(u), D_{ik}(u), n_{ik}(u)) \tag{2}$$

where $n_{ik}(u) = r_{ik}^S(u) + r_{ik}^P(u) + r_{ik}^D(u)$ and $S_{ik}(u) + P_{ik}(u) + D_{ik}(u) = 1$

Arbitrary hazard rate functions can be approximated with a set of discontinuous constant hazard rates over successive time intervals. The total follow-up time can be partitioned into M successive non-overlapping intervals indexed by $m = 1, ..., M$. We refer to interval m as U_m and write $u \in U_m$ to denote $u_m \leq u < u_{m+1}$. The length of U_m is $\Delta u_m = u_{m+1} - u_m$.

For each interval m, the proportions $S_{ik}(u)$, $P_{ik}(u)$, and $D_{ik}(u)$ are related to the time-varying hazards $h_{ik}^{SP}(u)$, $h_{ik}^{SD}(u)$, $h_{ik}^{PD}(u)$ according to a set of differential equations (**?**) expressed in Equation 3.

$$S_{ik}(u) = S_{ik}(u_m)e^{-(h_{ikm}^{SP}+h_{ikm}^{SD})(u-u_m)}$$

$$P_{ik}(u) = P_{ik}(u_m)e^{-h_{ikm}^{PD}(u-u_m)} + \frac{S(u_m)h_{ikm}^{SP}(e^{-(h_{ikm}^{SP}+h_{ikm}^{SD})(u-u_m)} - e^{-h_{ikm}^{PD}(u-u_m)})}{h_{ikm}^{PD} - h_{ik}^{SP} - h_{ikm}^{SD}} \tag{3}$$

$$D_{ik}(u) = 1 - S_{ik}(u) - P_{ik}(u)$$

In order to estimate h_{ikm}^{SP}, h_{ikm}^{SD}, and h_{ikm}^{PD} for each interval m, data at three time points were used: u_m, $u_{m+\frac{1}{2}}$, and u_{m+1}. The obtained estimates of the hazards for interval m were assigned to the time point corresponding to the mid point of the interval when used in the NMA.

For time points of a trial for which only OS data was reported (and no PFS), the number of patients that survived time interval m (r_{ikm}^{OS}) out of the number of patients alive at the beginning of interval (n_{ikm}^{OS}) were used for the model parameter estimation. A binomial likelihood was used for the conditional survival probability (p_{ikm}^{OS}) according to Equation 4.

$$r_{ikm}^{OS} \sim binomial(p_{ikm}^{OS}, n_{ikm}^{OS}) \tag{4}$$

The conditional survival probabilities inform $h_{ik}^{SD}(u)$ and $h_{ik}^{PD}(u)$ according to Equation 5.

$$p_{ikm}^{OS} = e^{-(h_{ikm}^{SD}+h_{ikm}^{PD})(u-u_m)} \tag{5}$$

where the obtained estimates of the hazards for interval m assigned to the time point corresponding to the mid point of the interval when used in the NMA.

8.1.1.3 Prior distributions

The prior distributions for the parameters of the model expressed with Equation 1 are displayed in Equation 6.

$$
\begin{pmatrix} \mu_{1i} \\ \mu_{2i} \\ \mu_{3i} \\ \mu_{4i} \\ \mu_{5i} \\ \mu_{6i} \end{pmatrix} \sim MVN \left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, T_\mu \right) \qquad T_\mu = \begin{pmatrix} 100 & 0 & 0 & 0 & 0 & 0 \\ 0 & 10 & 0 & 0 & 0 & 0 \\ 0 & 0 & 10 & 0 & 0 & 0 \\ 0 & 0 & 0 & 100 & 0 & 0 \\ 0 & 0 & 0 & 0 & 10 & 0 \\ 0 & 0 & 0 & 0 & 0 & 10 \end{pmatrix}
$$

(6)

$$
\begin{pmatrix} d_{1,1t} \\ d_{2,1t} \\ d_{3,1t} \end{pmatrix} \sim MVN \left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, T_d \right) \qquad T_d = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}
$$

$$
\sigma_{d_1} = uniform(0.2)
$$

The prior distribution for the probabilities at the beginning of each interval, u_m, is presented in Equation 7.

$$
(S_{ik}(u_m), P_{ik}(u_m), D_{ik}(u_m)) \sim Dirichlet(0, 0, 0)
$$

(7)

8.1.1.4 Model selection

The residual deviance and the deviance information criterion (DIC) wereused to compare the goodness-of-fit of the competing NMA models (**?**). DIC provides a measure of model fit that penalizes model complexity. In general, a more complex model results in a better fit to the data, demonstrating a smaller residual deviance. The model with the better trade-off between fit and parsimony has a lower DIC. A difference in the DIC of about 10 points was considered meaningful.

Figure 5: First line estimates of hazard ratios from stable to progression relative to gefitinib from the multi-state network meta-analysis

8.1.2 Absolute effects with 1L reference treatment

Absolute effects at 1L, the $\mu_{\cdot i}$, were estimated with a multi-state meta-analysis of gefitinib and the relative treatment effects, the $d_{\cdot,1t}$, were estimated using Equation 1. The simulated posterior distribution of the parameter estimates were then used to simulate a posterior distribution for PFS

and OS. Figure 7 reports mean estimates of PFS and OS by month for each 1L treatment and each fitted model.

Figure 6: First line estimates of progression-free survival and overall survival with gefitinib from multi-state meta-analyses
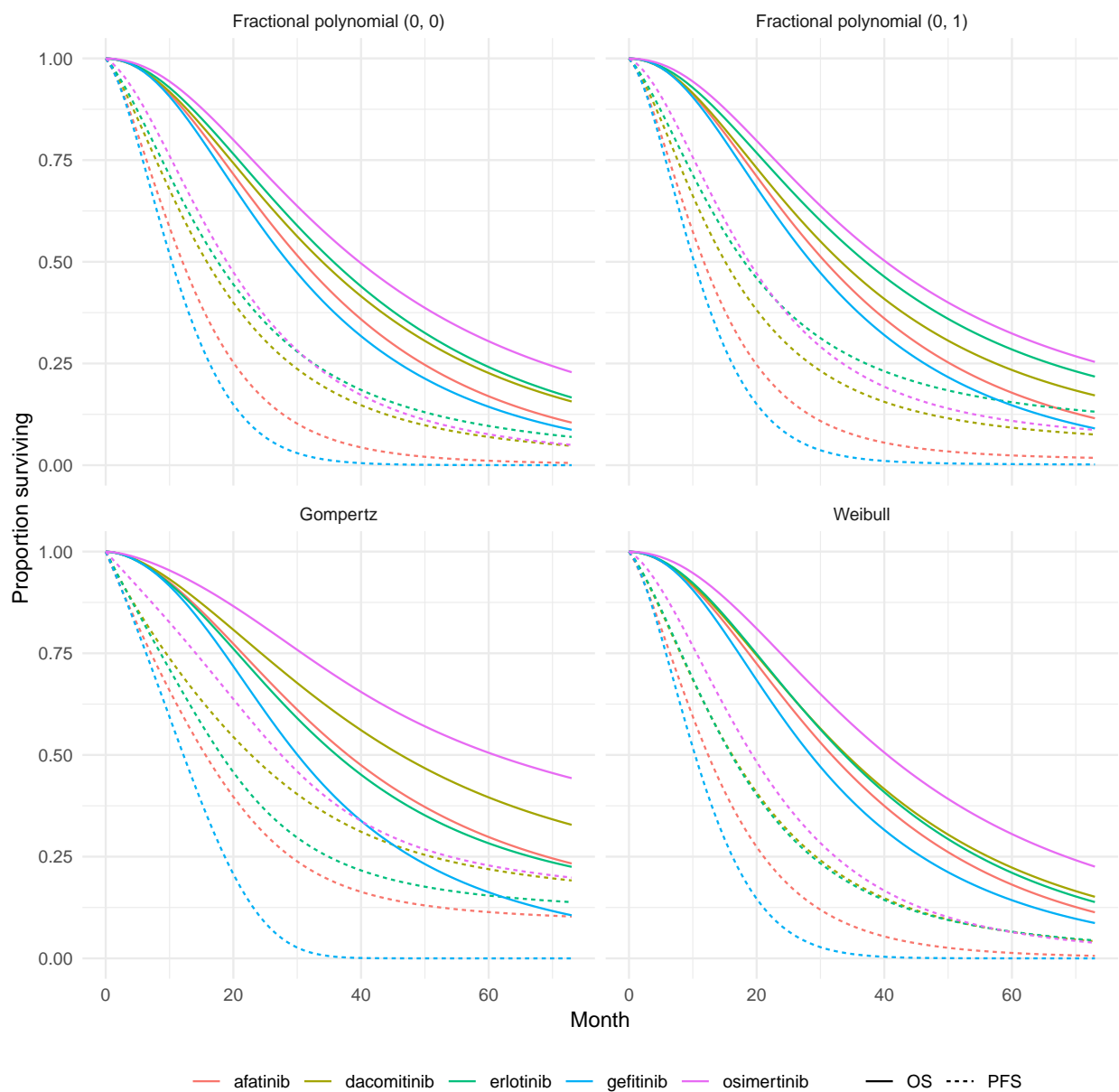
Figure 7: First line estimates of progression-free survival and overall survival for the competing interventions obtained with the multi-state network meta-analysis

Notes: The simulated posterior distribution of the parameters of the Bayesian multi-state NMA was used to simulate a distribution of progression-free survival and overall survival curves. The curves in the figure are posterior means. Curves with credible intervals are shown in the appendix.

8.1.3 Absolute effects in 2L

Recall that second line treatment depend on whether a patient has a T790M mutation. Separate 2L multi-state meta-analyses were consequently conducted to parameterize transitions among T790M positive and T790M negative patients. For T790M positive patients, we performed a meta-analysis of osimertinib. Among possible treatments for T790M negative patients, there was only trial data available for PBDC, so we performed a meta-analysis of PBDC (estimates for other 2L treatment options including PBDC + anti-VEGF therapy and 1st/2nd genration TKIs are assumed to be equivalent to PBDC in the economic model). Since T790M mutation status was not a prognostic factor or an effect modifier, the PBDC meta-analysis was conducted using an all-comer population. Mean estimates of PFS and OS by month for the 2L treatments are displayed in Figure 8a.

8.1.4 Use of meta-analysis results in economic model

Informed by model fit criteria, we used treatment effect estimates obtained with the fixed and random effects, Weibull and fractional polynomial models without a treatment effect for progression to death for the economic model. When a 3-state economic model is used, the 1L NMA is used to directly simulate disease progression. However, if a 4-state model is used, then both 1L and 2L estimates are incorporated. In particular, the 1L results are used to simulate the transition from S_1 to $P_1 \rightarrow S_2$ and the 2L results are used to simulate the transition from $P_1 \rightarrow S_2$ to P_2, P_2 to D, and $P_1 \rightarrow S_2$ to D (see Figure 2).

(a) **Platinum based doublet chemotherapy**
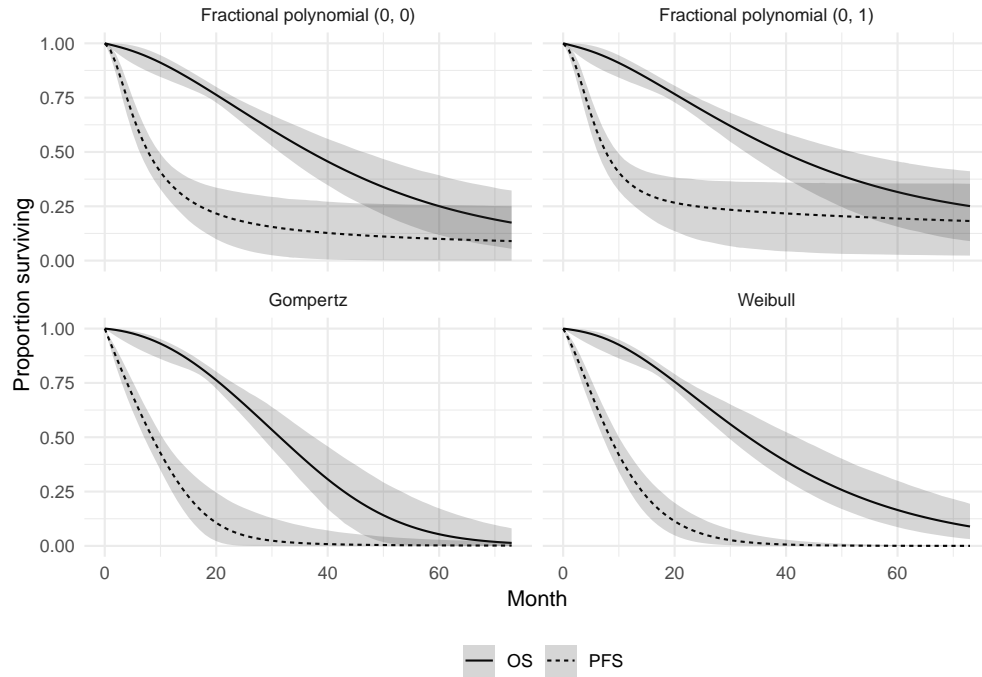
(b) **Osimertinib among T790M positive patients**

Figure 8: Second line estimates of progression-free survival and overall survival from the multi-state meta-analysis
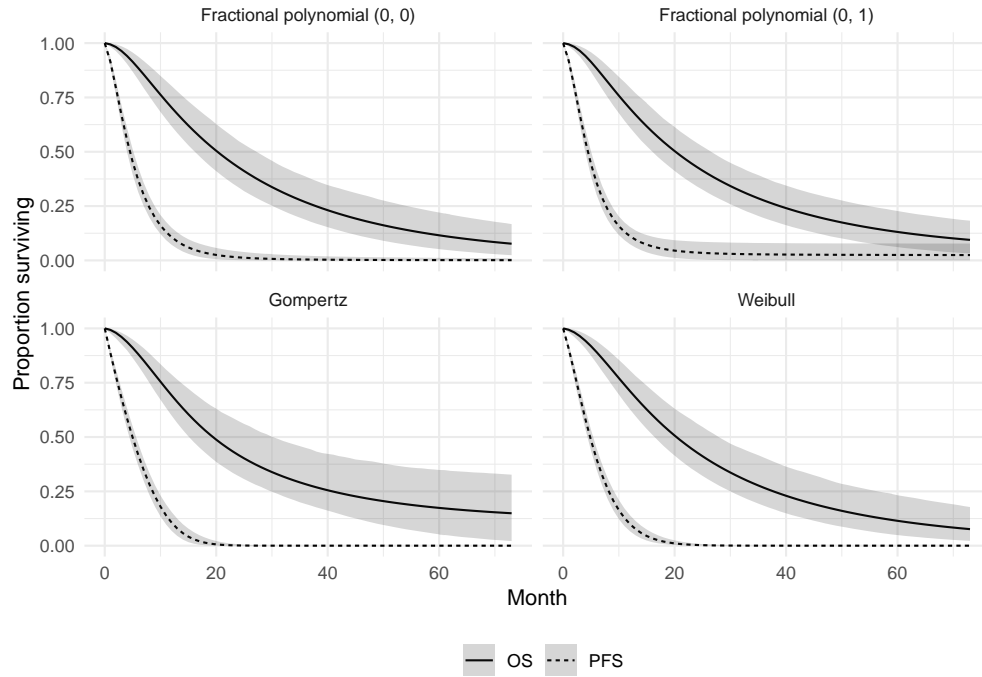
Notes: The simulated posterior distribution of the parameters of the Bayesian multi-state NMA was used to simulate a distribution of progression-free survival and overall survival curves. The solid lines are posterior means and the shaded region denotes the 95 percent credible interval.

8.2 Adverse events

The identified RCTs also provided the evidence base to estimate treatment specific adverse eent probabilities. Since adverse events in nearly all RCTs were reported as the number of patients experiencing the event, the NMA was performed on the proportion of patients experiencing the event of interest with a binomial likelihood and logit link (**?**, Chapter 2). Add details about pooling and the priors...

8.3 Utilities

Table 2: Utility by health state

Health state	Mean	Standard error	Reference
S1	0.7540	0.0000	?
P1/S2	0.6532	0.0222	?
P2	0.4734	0.0311	?

Table 3: Disutility due to adverse events

Adverse event	Mean	Standard error	Reference
Elevated alanine transaminase	0.0735	0.0185	?
Elevated aspartate transaminase	0.0735	0.0185	?
Diarrhea	0.0468	0.0155	?
Dry skin	0.0325	0.0117	?
Eye problems	0.0449	0.0148	?
Paronychia	0.0325	0.0117	?
Pneumonitis	0.0500	0.0120	?
Pruritus	0.0325	0.0117	?
Rash	0.0325	0.0117	?
Stomatitis	0.0325	0.0117	?

8.4 Health care sector costs

8.4.1 Treatment costs

Treatment costs are a function of drug acquisition and drug administration costs. Dosage for each drug based on Federal Drug Administration (FDA) labels is presented in Table 4. Patients using PBDC are, by default, assumed to use a combination of cisplatin and pemetrexed, although this can be adjusted in the **R** package. At a given treatment line, patients continue to take a drug until disease progression or the end of a treatment cycle. For example, a patient using erlotinib at 1L would take a 150 mg tablet each day until disease progression, at which point they would begin to use the 2L drug. Conversely, a patient using PBDC at 2L would discontinue cisplatin and pemetrexed after completing 6 21-day cycles.

In some cases, patients might be required to use multiple dosage forms to obtain the recommended dosage. In these cases, we assume that patients use the cheapest possible combination of dosage

Table 4: Drug dosage

Drug	Dosage
erlotinib	150 mg orally, once daily
gefitinib	250 mg orally, daily
afatinib	40 mg orally, once daily
dacomitinib	45 mg orally, once daily
osimertinib	80 mg orally, once daily
cisplatin	75mg/m2, 1x/cycle, 6 21-day cycles
pemetrexed	500 mg, 1x/cycle, 6 21-day cycles
bevacizumab	15mg/kg IV every 3 weeks with carboplatin/paclitaxel
nivolumab	240 mg every 2 weeks or 480 mg every 4 weeks
pembrolizumab	200 mg every 3 weeks
atezolizumab	1200 mg as an IV infusion over 60 min every 3 weeks

forms. For instance, dosage for nivolumab is 240 mg every 2 weeks, so patients are assumed to use 2 100mg vials and 1 40mg vials.

Dosing with cisplatin depends on body weight. The mean body surface area (BSA) in the United States is $1.6m^2$ for women and $1.9m^2$ for men, which implies a dose of of 120 mg for women of and 142.5 mg for men. The model assumes that there is no vial sharing, so patients use 1 100mg vial and 1 50 mg vial of cisplatin.

Wholesale acquisition costs (WACs) are shown in Table 5. Drug acquisition costs in the model are equal to the WACs adjusted for discounts and rebates. These discounts can be applied to uniquely to each drug in the **R** package, but are, by default, assumed to range from 20% to 30%. When historical data was available, we used the most recently available WAC from either ProspectoRx or AnalySource. There was no historical data for dacomitinib, so we used the publicly announced WAC of $12,400 per month.

Table 6 reports drug administration costs. Drug administration costs are based on US Current Procedures Terminology (CPT) codes and accrued each time a patient takes a drug.

8.4.2 Inpatient and outpatient costs

8.4.3 Adverse event costs

8.5 Productivity

Missed days of work to temporary disability = 151 (range = 120.8 - 181.2)

Reduction in hours worked due to permanent disability = 4 (range = 3 - 5)

8.6 Value of hope

To be added.

Table 5: Wholesale acquisition costs

Drug	Strength	Acquisition cost
erlotinib	150mg tablet	281.71
gefitinib	250mg tablet	257.06
afatinib	40mg tablet	235.05
dacomitinib	45mg tablet	407.39
osimertinib	80mg tablet	487.22
cisplatin	50mg/50ml vial	14.51
cisplatin	100mg/100ml vial	35.00
cisplatin	200mg/200ml vial	100.00
pemetrexed	100mg vial	676.52
pemetrexed	500mg vial	3382.60
bevacizumab	100mg/4ml vial	199.24
bevacizumab	400mg/16ml vial	199.24
nivolumab	100mg/10ml vial	262.21
nivolumab	40mg/4ml vial	262.21
pembrolizumab	100mg/4ml vial	1162.41
atezolizumab	1200mg/20ml vial	444.03

Table 6: Drug administration costs

Drug	Administration cost	Source
erlotinib	0.00	N/A
gefitinib	0.00	N/A
afatinib	0.00	N/A
dacomitinib	0.00	N/A
osimertinib	0.00	N/A
cisplatin	91.72	CPT 96417, 96415
pemetrexed	136.15	CPT 96413
bevacizumab	91.72	CPT 96413, 96415, 96417
nivolumab	136.15	CPT 96413
pembrolizumab	136.15	CPT 96413
atezolizumab	136.15	CPT 96413

Table 7: Inpatient medical costs

Health state	Mean	Standard error	Reference
S1	7,699	0	None
P1/S2	7,699	0	?
P2	12,577	0	?

Table 8: Outpatient medical costs

Health state	Mean	Standard error	Reference
S1	127,860	0	None
P1/S2	127,860	0	?
P2	127,728	0	?

Table 9: Costs of adverse events

Adverse event	Mean	Lower	Upper	Reference
Diarrhea	6,462	5,169	7,754	DRG 391
Dry skin	940	752	1,128	?
Elevated alanine transaminase	3,900	3,120	4,680	?
Elevated aspartate transaminase	3,900	3,120	4,680	?
Eye problems	2,737	2,189	3,284	?
Paronychia	7,788	6,230	9,345	DRG 602
Pneumonitis	7,728	6,182	9,273	DRG 193
Pruritus	1,184	947	1,420	?
Rash	940	752	1,128	?
Stomatitis	8,101	6,480	9,721	DRG 157

Table 10: Weekly wages by gender and employment status

Gender	Employment status	Percentage	Weekly wage
Female	Full-time	72.4%	$796
Female	Part-time	23.3%	$326
Female	Unemployed	4.3%	$0
Male	Full-time	84.1%	$973
Male	Part-time	11.5%	$321
Male	Unemployed	4.4%	$0

9 Simulation and uncertainty analysis

9.1 Parameter uncertainty

Parameter uncertainty is quantified using probabilistic sensitivity analysis (PSA), which propagates uncertainty in the model input parameters throughout the model by randomly sampling the input parameters from suitable probability distributions (**?**), (**?**).Probability distributions are determined according to the distributional properties of the statistical estimates, which, in turn, depend on the statistical techniques used, and the distributions of the underlying data. Table 11 displays the probability distribution that we used for the model parameters. When conducting a CEA, the results of the PSA are summarized using standard measures from the literature including cost-effectiveness planes, cost-effectiveness acceptability curves, and estimates of the expected value of perfect information (**?**), (**?**), (**?**). Furthermore, all point estimates from a CEA or MCDA are reported along with 95 percent confidence intervals.

Table 11: Probability distributions for probabilistic sensitivity analysis

Parameter(s)	Distribution
Health state transitions	
Bayesian multi-state NMA	Simulated posterior distribution
Adverse events	
Bayesian NMA for adverse events	Simulated posterior distribution
Utility	
Health state utility	Normal
Adverse event disutilities	Normal
Health care sector costs	
Drug acquisition and administration cost	Fixed
Discounts and rebates applied to drug acquisition cost	Uniform
Inpatient costs	Gamma
Outpatient costs	Gamma
Adverse event costs	Normal
Productivity costs	
Missed work days	Uniform
Reduction in hours worked	Uniform

9.2 Structural uncertainty

Health state probabilities are typically very sensitive to the parametric distributions used in a survival models. We explore this sensitivity by providing options to choose from a number of different distributions. As discussed in **??** these included the Weibull, Gompertz, and two second order fractional polynomial models $\{(p_0, p_1) = (0,0), (0,1)\}$. Furthermore, both 4-state (Figure 2) and 3-state (Figure 3) model structures can be used. [Need to add detail.]

9.3 Implementation

The individual-level CTSTM was implemented using the R package **hesim** developed by IVI. The package provides a general framework for integrating parameter estimates from a statistical model with different types of simulation models for economic evaluation. Parameter uncertainty is propagated throughout *hesim* models using PSA. Specifically, we randomly drew all input parameters as described in Section 9.1 and simulated the CTSTM for each randomly sampled parameter set. All computationally intensive steps in **hesim** simulation models are written in C++ so the simulations are fast enough to be run with a PSA in web applications.

Appendices

A Systematic Literature Review

A.1 Treatment effects for transition rates

Details regarding eligibility criteria defining the scope of the studies considered relevant are outlined in Table A1, Table A2, and Table A3.

A.2 Utilities

Health state utility values (EQ-5d, HUI2, HUI-3, SF-6D) for the different health states were used in the model as well as disutility estimates associated with adverse events identified by focusing on review or overview studies. Furthermore, published mapping algorithms that allow a non-preference-based measure (generic or disease-specific measure) to be mapped onto a generic preference-based measure are of interest, as well as mapping algorithms between different generic preference-based health state utility values (e.g. between SF-6D and EQ5D) will be searched for. See Appendix Table 4 for study selection criteria.

A.3 Resource use, productivity, and cost

Relevant evidence regarding resource use, productivity and cost estimates were identified by means of a review of published CEA modelling studies in NSCLC relevant for the US setting. See Appendix Table 5 for study selection criteria.

A.4 Study identification

Relevant studies were identified by searching the following databases using predefined search strategies: Medical Literature Analysis and Retrieval System Online (MEDLINE), Excerpta Medica

Table A1: PICOS criteria for review of treatment effects (metastatic 1L population)

PICOS	Criteria
Population	Adult patients with metastatic non-squamous NSCLC who are EGFR positive and without prior treatment for their disease.
Interventions	The following drugs as monotherapy or in combination with other drugs

- erlotinib
- afatinib
- gefitinib
- osimertinib
- dacomitinib

Comparators	

- placebo
- best supportive care (BSC), defined as whichever therapy was judged to be appropriate by the treating physician.
- any intervention of interest
- any treatment that facilitates an indirect comparison

Outcomes	

- OS
- PFS
- TTP

Study design	RCTs
Other	English language

Note: Trials where the overall study population is an all-comer population, but subgroup results are reported for the target population of interest are included. A trial or subgroup where the study population is a mixture of non-squamous and squamous patients is included if over 90% is non-squamous.

database (EMBASE), and Cochrane Central Register of Controlled Trials. The study design filters recommended by the Scottish Intercollegiate Guidelines Network (SIGN) for MEDLINE and EM-BASE were used to identify clinical trials. The search included terms related to the generic and brand name of the interventions of interest. The US National Institutes of Health Clinical Trial Registry and EU Clinical Trials Registry were also searched to identify completed clinical trials

not yet published to identify any completed or ongoing trials that meet the criteria with results available.

For utility studies, and CEA studies providing evidence on healthcare utilization and costs, the following additional databases were searched in addition: NHS Economic Evaluation Database (NHS EED), Health Economic Evaluations Database (HEED), THe Health Economics Research Centre-maintained mapping algorithm database, and The University of Sheffield's ScHARRHUD database of health utilities' evidence.

Where there were gaps in the identified evidence, the search was expanded to include the grey literature, such as reports published by NICE.

A.5   Study selection

Two reviewers, working independently, reviewed all abstracts and proceedings identified in each of the searches according to the selection criteria, with the exception of outcome criteria in the efficacy and safety searches, which were only applied during the screening of full-text publications. All studies identified as eligible studies during abstract screening were then screened at a full-text stage by the same two reviewers. The full-text studies identified at this stage were included for data extraction. Following reconciliation between the two investigators, a third reviewer was included to reach consensus for any remaining discrepancies. The process of study identification and selection is summarized with Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagrams (including reasons for exclusion at both the abstract and full text screening stage).

A.6   Data collection

For the clinical studies, two reviewers, working independently, extracted data on study characteristics, interventions, patient characteristics, and outcomes for the final list of included studies. Following reconciliation between the two reviewers, a third reviewer was included to reach consensus on any remaining discrepancies. For all outcomes of interest, information regarding point estimates, variability and uncertainty was obtained. For PFS and OS, hazard ratios and associated information regarding uncertainty were extracted. If results were reported as forest plots, the point estimate and 95$ percent confidence interval were extracted using DigitizeIt software version 2.1.4 (Bormisoft - Informer Technologies, Inc.). Kaplan Meier curves will also be digitzed using DigitizeIt and the proportion of patients free of the event over time will be extracted and the number of patients at risk over time. Adverse events were collected, if reported. Data collected included number and percent of patients with: any grade 3 or 4 adverse event, any grade 3 or 4 treatment related adverse event, any serious adverse event, any treatment related serious adverse event, and death within 30 days of last treatment. The duration over which safety outcomes were reported were also captured. Discontinuation due to adverse events were also examined.

For utility and cost-effectiveness studies, relevant information for the model was extracted from the source publications and checked by a second-reviewer.

A.7   Limitations

Despite the strengths of the proposed SLR, some limitations are applicable to all SLRs that should be acknowledged. As the evidence base is continually growing, any trials published after the search date will not be captured. Further, any trials that are published close to the search date but are not yet indexed in the databases at the time of the search will not be captured by the search of MEDLINE, EMBASE, and Cochrane Central Register of Controlled Trials. As with any literature

review, this SLR is limited by the use of published data. There is a risk of publication bias as some clinical studies fail to be published while others are published only in abstract form, which presents limited information. Finally, the search and selection were restricted to trials published in English.

Table A2: PICOS criteria for review of treatment effects (metastatic 2L population)

PICOS	Criteria
Population	Adult patients with metastatic non-squamous NSCLC who are EGFR+ positive and who have experienced progression after one line of prior treatment.
Interventions	The following drugs as monotherapy or in combination with other drugs • erlotinib • afatinib • gefitinib • osimertinib • dacomitinib • nivolumab • pembrolizumab • atezolizumab • bevacizumab • platinum-based doublet therapy
Comparators	• placebo • best supportive care (BSC), defined as whichever therapy was judged to be appropriate by the treating physician. • any intervention of interest • any treatment that facilitates an indirect comparison
Outcomes	• OS • PFS • TTP
Study design	RCTs
Other	English language

Note: Trials where the overall study population is an all-comer population, but subgroup results are reported for the target population of interest are included. A trial or subgroup where the study population is a mixture of non-squamous and squamous patients is included if over 90% is non-squamous.

Table A3: PICOS criteria for review of treatment effects (metastatic 2L+ population)

PICOS	Criteria
Population	Adult patients with metastatic non-squamous NSCLC who are EGFR+ positive and who have experienced progression after two or more prior prior treatments.
Interventions	The following drugs as monotherapy or in combination with other drugs

- nivolumab
- pembrolizumab
- atezolizumab
- bevacizumab
- platinum-based doublet therapy

Comparators

- placebo
- best supportive care (BSC), defined as whichever therapy was judged to be appropriate by the treating physician.
- any intervention of interest
- any treatment that facilitates an indirect comparison

Outcomes

- OS
- PFS
- TTP

Study design	RCTs
Other	English language

Note: Trials where the overall study population is an all-comer population, but subgroup results are reported for the target population of interest are included. A trial or subgroup where the study population is a mixture of non-squamous and squamous patients is included if over 90% is non-squamous.

B Network meta-analysis

B.1 Feasibility assessment

With an NMA, interest centers on the comparison of the treatment effects of interventions that have not been studied in a head-to-head fashion. In order to ensure that these indirect comparisons of interventions are not affected by differences in study effects (i.e. known and unknown prognostic factors) between studies, we only considered the treatment effects of each trial. This consideration implied that all interventions indirectly compared have to be part of one network of trials where each trial has at least one intervention (active or placebo) in common with another trial. For interventions of interest that were not part of the same network, then it was not possible to perform an indirect comparison of treatment effects of these interventions without a substantial risk of bias (**??**).

In combining direct and indirect evidence in an NMA, trials must be reasonably similar. Patients are randomized only within trials, not across trials, so there is a risk that patients participating in different trials differ with respect to demographic, disease or other characteristics. In addition, features of the trials themselves may differ. If these trial or patient characteristics are effect modifiers, i.e. they affect the treatment effects of an intervention versus a control, then there are systematic differences in treatment effects across trials. Systematic differences in known and unknown effect-modifiers among studies comparing the same interventions in direct fashion result in between-study heterogeneity. An imbalance in the distribution of effect modifiers between studies comparing different interventions will result in transitivity or consistency violations and therefore biased indirect comparisons (**??**).

In order to gauge the appropriateness of proceeding with an NMA, IVI's feasibility assessment included: (i) an assessment of whether the RCT evidence for the interventions of interest formed one evidence network for each outcome by line of treatment conditional upon prior treatment (class); and (ii) an assessment of the distribution of study and patient characteristics that may have had relative treatment effects across the direct comparisons of the evidence networks.

B.2 Evaluation of consistency between direct and indirect comparisons

Prior to the actual NMA, the consistency between direct and indirect comparisons were evaluated for networks that include closed loops. For comparisons (i.e., contrasts) that are part of a closed loop made up of more than 1 RCT connecting different interventions, we assessed the consistency by comparing the relative treatment effect for one contrast of this loop based on direct information with the corresponding estimates based on indirect information (**?**).

B.3 Estimation of the relative treatment effects under the assumption of consistency

Based on the findings of the feasibility assessment, the results of the RCTs that are part of one evidence network and deemed sufficiently similar were synthesized by means of NMAs by outcome of interest. Under the assumption of consistency, the NMA model relates the data from the individual studies to basic parameters reflecting the (pooled) relative treatment effect of each intervention compared to an overall treatment of reference. The NMA models were extended by incorporating covariates (i.e., meta-regression analysis) only if there was an imbalance in treatment potential effect modifiers across trials in the network (**?**).

B.4 Software

The parameters of the different models were estimated using a Markov Chain Monte Carlo (MCMC) method implemented in the JAGS software package. All JAGS analyses were run using R statistical software (**?**).