Study Protocol

Open-Source Value Project - Sequential treatment strategies for patients with metastatic EGFR+ non-small cell lung cancer

Version 2

August 17, 2018

Administrative structure

Innovation and Value Initiative	Jennifer Bright, MPA; Executive Director
	Jason Shafrin, PhD; Director of Research
	11100 Santa Monica Boulevard, Suite 500
	Los Angeles, CA 90025
	Phone: 310 984 7793
	info@thevalueinitiative.org
Project team	Jeroen P Jansen, PhD
	Devin Incerti, PhD
	Ina Zhang, MPH
	Shivani Mehta, BSc
	Florence Wilson, MSc
	Prodyumna Goutam, PhD

Version History

Version	Changes	Date issued
1	--	July 1, 2018
2 (public release)	Revision based on questions by IVI Strategic Advisory Panel (SAP) before public release	August 17, 2018

Table of Contents

List of Tables

List of Figures

Abbreviations

1L	First line
2L	Second line
2L+	Beyond second line
BSC	Best supportive care
CEA	Cost-effectiveness analysis
CENTRAL	Cochrane Central Register of Controlled Trials
CR	Complete response
CTSTM	Continuous-time state transition model
DIC	Deviance information criterion
DOR	Duration of response
ECOG	Eastern Cooperative Oncology Group
EGFR	Epidermal Growth Factor Receptor
EMBASE	Excerpta Medica dataBASE
HEED	Health Economic Evaluations Database
IVI	Innovation and Value Initiative
KM	Kaplan-Meier
MCDA	Multi-criteria Decision Analysis
MCMC	Markov Chain Monte Carlo
MEDLINE	Medical Literature Analysis and Retrieval System Online
NHS EED	National Health Service Economic Evaluation Database
NICE	National Institute for Health and Care Excellence
NMA	Network meta-analysis
NMB	Net-monetary benefit
NSCLC	Non-small cell lung cancer
ORR	Overall response rate
OS	Overall survival
OSVP	Open-Source Value Project
PFS	Progression free survival
PICOS	Population, Interventions, Comparisons, Outcomes, and Study Design
PR	Partial response
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PSA	Probabilistic sensitivity analysis
QALY	Quality adjusted life year
RCT	Randomized controlled trial
SIGN	Scottish Intercollegiate Guidelines Network
SLR	Systematic Literature Review
TKI	Tyrosine kinase inhibitor
TMB	Total mutation burden
TTP	Time to progression

1. Introduction

1.1. The open-source value project (OSVP)

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 States' 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.

1.2. Topic definition

IVI will conduct its next OSVP within oncology. When selecting the tumor type of interest, IVI considers criteria such as burden of disease, development of innovative treatment, 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.(1) Non-small cell lung cancer (NSCLC) accounts for an estimated 85% of lung cancer cases and comprises adenocarcinoma, squamous cell carcinoma, and large cell carcinoma.(2) The five-year survival of stage IV NSCLC is less than 2%.(3) Given the rapid pace of development in NSCLC, the scope of the OSVP needs to be limited to a very specific sub-population in order to ensure a model can be developed in a reasonable time frame that allows IVI to demonstrate the typical areas of uncertainty in value assessment in oncology. The selected target population of interest for the current 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.(4). 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.

2. Objective

The aim is 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. Purpose

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. Potential sequential treatment strategies of interest are outlined in **Figure 1**. The final treatment strategies – starting with 1st line treatment (1L), followed by 2nd line treatment (2L), and treatment beyond 2nd line (2L+) - that can be evaluated with version 1 of the model will be informed by the available evidence base, guidelines, and clinician input. The model is primarily designed to evaluate sequences starting with 1L, but sequences starting with 2L may be analyzed given a certain 1L treatment received. If the evidence base is too limited, the model may only allow comparing sequences starting with 1L though.

Figure 1. Potential sequential treatment strategies of interest to be compared with the model, depending on the available evidence

Metastatic EGFR+ Non-Small Cell Lung Cancer
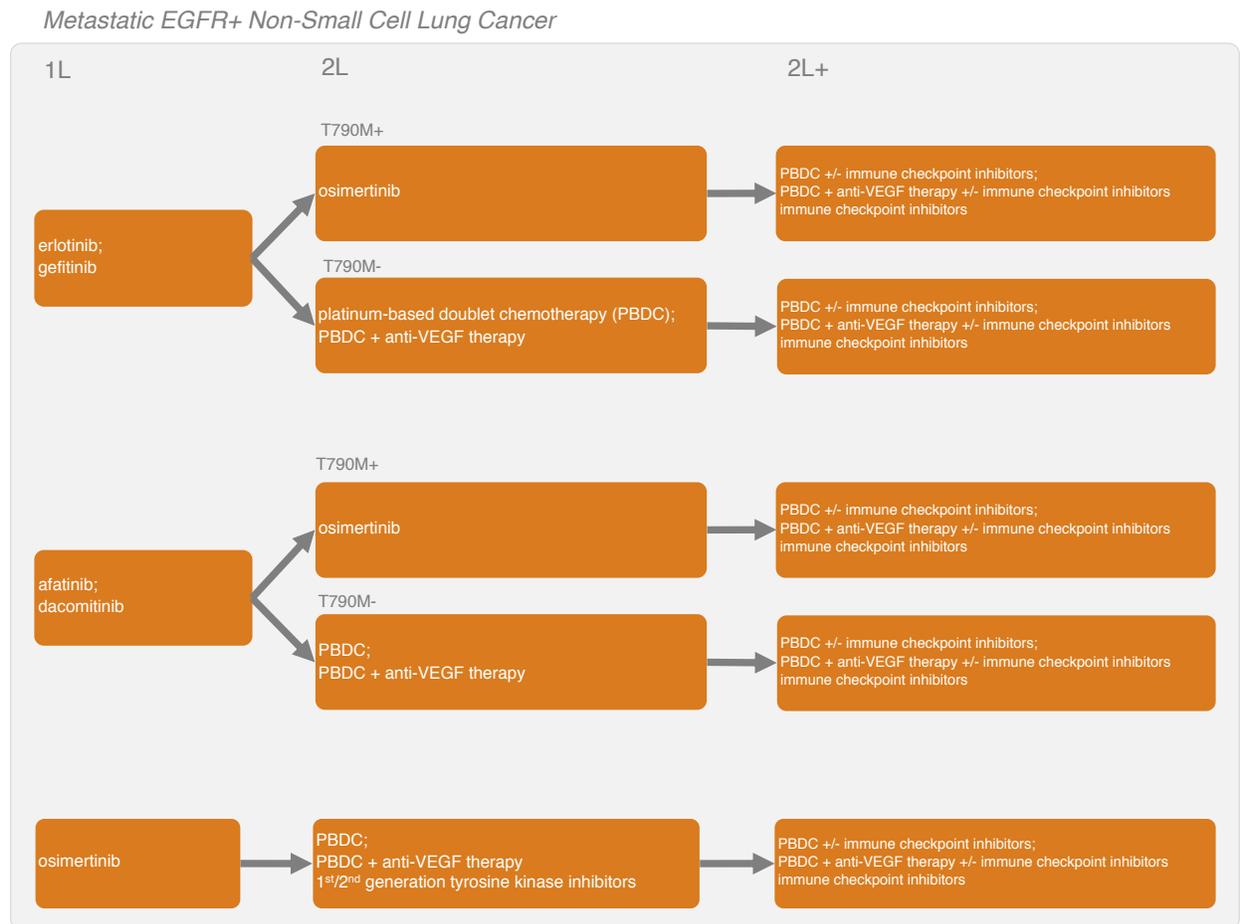
The IVI-NSCLC(egfr+) model is suitable for informing decisions for specific (sub)populations, but is not suitable for making predictions at the individual level, nor should it replace the patient-physician shared decision-making process. Local decision makers can modify the model to perform analysis 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.(5)

The MCDA will be implemented according to Thokala et al. and based on the following criteria: progression free survival (PFS), overall survival (OS), adverse events, expected time on first, second, and third line 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.(6) 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.

Garrison, Kamal-Bahl, and Towse 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.(7)

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.(7) This element of value revolves around the risk that a particular treatment may or may not work for a particular patient. 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.

Conventional value assessments focus on value to the sick, but recent research provides a framework for valuing medical technology for the healthy (i.e., "insurance value") as well.(8) The availability of an efficacious intervention for a certain disease state provides some degree of protection against the physical and financial risk among healthy individuals at risk for the disease. The insurance value framework can be implemented with knowledge of two additional parameters beyond those in conventional cost-effectiveness analysis: the probability of illness, and the marginal rate of substitution between the sick and the well states.

The IVI-NSCLC(egfr+) model allows users to incorporate value of hope and insurance value into their analyses, while noting that these approaches are less well-established than conventional cost-effectiveness analysis. Future versions of the IVI-NSCLC(egfr+) may consider incorporating other value components, such as option 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 to be available in the public domain:

Source code: R and C++ code for the model will be 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.

Model Interface: For users not be well-versed in the programming language, a web application for running the model online will be 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 will be developed allowing those who are not experts in modeling or health economics, to interact with the model.

6.   Patient input

The OSVP model development process will involve stakeholder feedback, in particular from patients and their representatives. In parallel to the development of version 1 of the IVI-NSCLC(egfr+) model, as described in this document, patient focus groups are being conducted to investigate how NSCLC patients value various treatment attributes such as efficacy, safety, affordability and convenience and how they prioritize these in their treatment decisions. After the focus groups are completed in the summer of 2018, IVI will be conducting detailed patient surveys to measure the value of different treatment attributes and health states as part of the Patient Experience Study. For version 1 of the IVI-NSCLC(egfr+) model, the value of different treatment attributes beyond efficacy and safety (e.g. mode of administration, and number of years a treatment has been available after regulatory approval) will be incorporated as part of the MCDA. Findings of the focus groups and Patient Experience Study that may have a direct impact on model structure and evidence synthesis will be incorporated in future iterations of the IVI-NSCLC(egfr+) model, building on the quantitative framework version 1 provides. Additional detail is provided in an accompanying document (**Stakeholder Input Process**).

7.   Methodology

7.1.   Model structure

7.1.1.   Disease model

The IVI-NSCLC(egfr+) 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 assume 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) (**See Section 7.4.2**). In the individual-level model, time to each state r that can be entered from state q is sampled based on the 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.

Figure 2: Model structure with 5 states describing development of disease over time for a treatment sequence starting with 1L, followed by 2L, and finally 2L+ treatments
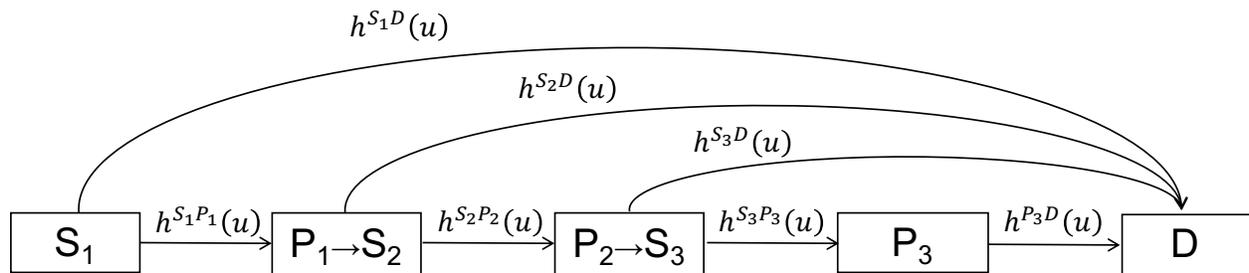
S_1 = Progression-free (stable disease) with 1L treatment

P_1 = Progression with 1L treatment

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

P_2 = Progression with 2L treatment

S_3 = Progression-free (stable disease) with 2L+ treatment

P_3 = Progression with 2L+ treatment

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

$h^{S_3 P_3}(u)$ = hazard for transitioning from progression-free to progression with 2L+ treatment at time u

$h^{S_3 D}(u)$ = hazard for transitioning from progression-free to dead with 2L+ treatment at time u

$h^{P_3 D}(u)$ = hazard for transitioning from progression to dead with 2L+ treatment at time u

If the evidence base is too limited to explicitly model the transition rates by line of treatment, the sequence of 1L, 2L, and 2L+ will be represented with a simplified model as depicted in **Figure 3** and **Figure 4.** For example, if for a sequence where a 1st generation tyrosine-kinase inhibitor (TKI) is used in the 1L setting followed by osimertinib due to T790-mutation related resistance in the 2nd line and there are no studies reporting PFS and OS for subsequent treatments in the 2L+ setting, then the model structure for that sequence will be in line with **Figure 3**. Or, if for a treatment sequence starting with 1L osimertinib and there are no studies reporting PFS and OS with 2L and 2L+ treatment representative of an osimertinib experienced patient population, then the model structure for that sequence will be in line with Figure 4, where $h^{PD}(u)$ includes 2L and 2L+ treatment.

Figure 3: Simplified 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
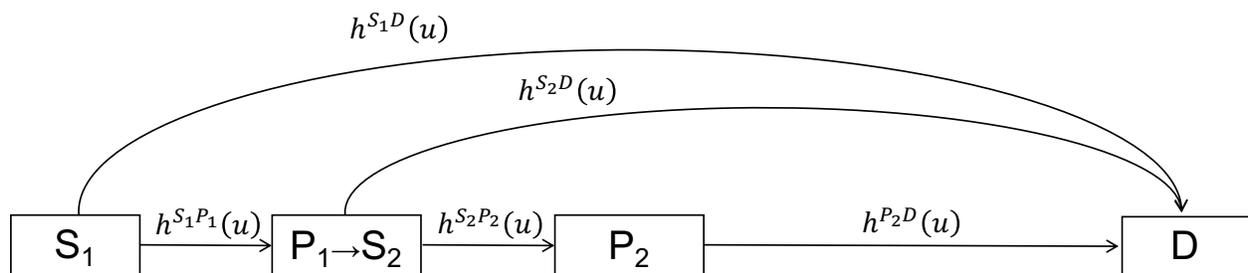
S_1= Progression-free (stable disease) with 1L treatment

P_1= Progression with 1L treatment

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

$h^{S_1P_1}(u)$= 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

$h^{S_2P_2}(u)$= hazard for transitioning from progression-free to progression with 2L treatment at time u

$h^{S_2D}(u)$= hazard for transitioning from progression-free to dead with 2L treatment at time u

$h^{P_2D}(u)$= hazard for transitioning from progression on 2L to dead at time u

Figure 4: Simplified 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
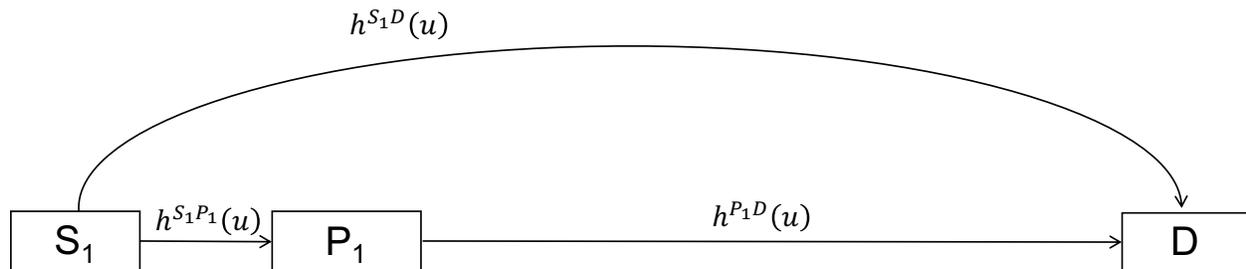
S_1 = Progression-free (stable disease) with 1L treatment

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

$h^{S_1 P_1}(u)$ = 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

$h^{P_1 D}(u)$ = hazard for transitioning from progression on 1L to dead at time u

7.1.2. Adverse events

The model will include 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.

7.1.3. Cost and utility

Each 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.). Unless we find strong evidence to the contrary, we will assume 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.

Drug acquisition and administration related costs will be 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.

Value of hope will be 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.

7.1.4. 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 can model 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 $g^{-1}(\cdot)$ is an inverse transformation function. However, the extent heterogeneity can be modeled depends on the level of detail in the available evidence.

7.1.5. 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 reset" approach is used, then health state probabilities can be calculated analytically using the Aalen-Johansen 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.(9-12) In our case, the multi-state NMA is a mixture between the two: each time a patient begins a new treatment time resets since separate multi-state NMA's will be conducted by treatment line, but time does not reset within a given treatment line since hazard rates depend on time since treatment initiation conditional on treatment line in the multi-state NMA. To accommodate a mixture of "clock reset" and "clock forward" approaches, we use an individual-level model.

7.2. Model output

The model will simulate 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.

Table 1. Model outcomes

Health outcomes	Health state probabilities, survival curves, life expectancy, life-years, quality-adjusted life-years
Risks	Adverse events (e.g., fatigue, hair loss, nausea/vomiting, diarrhea, neutropenia)
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	CEA-based NMB, MCDA-based "total value" score

7.2.1. 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 will be 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 will be 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 will also be calculated using a default discount rate of 3%.

7.2.2. Risks

The model will calculate 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 will be 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 will be reported.

7.2.3. Costs

Costs will be calculated for multiple categories and separated into health care sector and non-heath care sector costs as recommended by the Second Panel on Cost-Effectiveness in Health and Medicine.(13) We expect health care sector cost categories to include 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 will include productivity losses. The intent is to use both a human-capital approach and friction-cost approach to estimate productivity losses associated with the different treatment

sequences. Its feasibility and level of detail regarding productivity related cost estimates that will be incorporated depends on the available data in the literature.

7.2.4. Value assessment

Value will be assessed using two frameworks: CEA and MCDA as described in **Section 3.1**. If CEA is used for value assessment, then the value of treatment is estimated using the NMB, CEA from a societal perspective will include productivity losses while analyses from a health care sector perspective will 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 .167, .33, and .5, respectively. Third, to aggregate results, we multiply 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.

7.3. Simulation and uncertainty analysis

7.3.1. 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 their joint probability distribution.(14,15) 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 2** displays the probability distribution that we expect to use for the model parameters. When conducting a CEA, the results of the PSA will be summarized using standard measures from the literature including cost-effectiveness planes, cost-effectiveness acceptability curves, and estimates of the expected value of perfect information.(16-19) Furthermore, all point estimates from a CEA or MCDA will be reported along with 95% credible intervals.

Table 2. Probability distributions for probabilistic sensitivity analysis

Parameter estimates from Bayesian multi-state NMA (efficacy and safety)	Simulated posterior distribution generated from Markov Chain Monte Carlo (MCMC) algorithm
Health state utility	Beta distribution
Resource use and costs	Gamma distribution

7.3.2. Structural uncertainty

Health state probabilities are typically very sensitive to the parametric distributions used in a multi-state statistical model. We explore this sensitivity by providing options to choose from a number of different distributions (e.g., exponential, Weibull, Gompertz, log-logistic, and fractional polynomial distributions). Furthermore, formal Bayesian model averaging may be considered to generate point estimates derived from weighted averages of models fit using different distributions with weights based on posterior probabilities.(20)

In addition, as noted by the National Institute for Health Care Excellence (NICE) Decision Support Unit, methods for extrapolating outcomes beyond the time horizons of the original studies can have a large impact on results.(21) Moreover, the hazard rates and treatment effects on the hazard rate estimated using trial data may not generalize to the extrapolation period. We will therefore run a number of scenario analyses to quantify the impact of different assumptions during the extrapolation period on results. These will include:

- Modeling health state transitions with different survival distributions (e.g., exponential, Weibull, Gompertz, log-logistic, and fractional polynomial distributions) during the extrapolation period;

- Modeling three separate treatment effect scenarios: (i) no difference in treatment effects across competing interventions, (ii) treatment effects are the same as during the pre-extrapolation period, and (iii) treatment effects diminish toward zero.

Finally, different methods for simulating adverse events, costs, and utilities will be considered if a review of the literature suggests that multiple approaches are scientifically valid.

7.3.3. Implementation

The individual-level CTSTM will be 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. In our case, we randomly draw all input parameters from appropriate probability

distributions as described in **Section 7.3.1,** and simulate 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.

Although *hesim* supports standard individual-level CTSTMs using R alone, the package also provides a comprehensive C++ library for developing custom simulations with C++. We may take this approach by customizing *hesim*'s individual-level CTSTM classes and adding new features such as functions for simulating adverse events.

7.4. Source data and parameter estimation

Key parameters for the model relate to: (i) the treatment effects of the interventions used for the different lines of treatment in terms of PFS and OS; (ii) utilities; (iii) healthcare resource use; and (iv) productivity. Estimates for these parameters will be based on currently available published evidence identified by means of a systematic literature review (SLR) and synthesized with meta-analysis techniques where appropriate.

7.4.1. Systematic literature review

7.4.1.1. Treatment effects

The systematic literature reviews of treatment effects will focus on randomized controlled trials (RCTs) evaluating the efficacy of relevant competing interventions for the treatment of metastatic EGFR+ non-squamous NSCLC by line of treatment, i.e. 1L, 2L, and 2L+. Primary outcomes of interest are OS, PFS and time to progression (TTP). If the evidence base is too limited regarding PFS, TTP and OS estimates, the scope will be expanded by including the following secondary outcomes: response, duration of response (DOR), EORCT QLC-C30, and EORTC-LC 30. Details regarding eligibility criteria defines the scope of the studies considered relevant are outlined in **Tables 3-5 in Appendix A**. The identified evidence will be used to estimate the relative treatment effects for each intervention versus a defined intervention of reference by line of treatment conditional upon the prior treatment (class), as well as to estimate the absolute treatment effects with the corresponding reference treatments.

7.4.1.2. Utilities

Health state utility values (EQ-5d, HUI2, HUI-3, SF-6D) for the different health states used in the model as well as disutility estimates associated with adverse events will be identified by focusing on review or overview studies. If these are not available, the focus will include primary publications. Furthermore, published mapping algorithms that would 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 **Table 6 in Appendix A** for study selection criteria.

7.4.1.3. Resource use, productivity, and cost

Relevant evidence regarding resource use, productivity and cost estimates will be identified by means of a review of published CEA modelling studies in NSCLC relevant for the US setting. See **Table 7 in Appendix A** for study selection criteria.

7.4.1.4. Study identification

Relevant studies will be identified by searching the following databases using predefined search strategies: Medical Literature Analysis and Retrieval System Online (MEDLINE), Excerpta Medica database (EMBASE), and Cochrane Central Register of Controlled Trials. The study design filters recommended by the Scottish Intercollegiate Guidelines Network (SIGN) for MEDLINE and EMBASE will be used to identify clinical trials (http://www.sign.ac.uk/search-filters.html). The search will include terms related to the generic and brand name of the interventions of interest (**Appendix A**). The US National Institutes of Health Clinical Trial Registry (http://www.clinicaltrials.gov) and EU Clinical Trials Registry (https://www.clinicaltrialsregister.eu) will also be 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 will be 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.

When there are gaps in the identified evidence, the search may be expanded to include the grey literature, such as reports published by NICE.

7.4.1.5. Study selection

Two reviewers, working independently, will review 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 will only be applied during the screening of full-text publications. All studies identified as eligible studies during abstract screening will then be screened at a full-text stage by the same two reviewers. The full-text studies identified at this stage will be included for data extraction. Following reconciliation between the two investigators, a third reviewer will be included to reach consensus for any remaining discrepancies. The process of study identification and selection will be 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).

7.4.1.6. Data collection

For the clinical studies, two reviewers, working independently, will extract data on study characteristics, interventions, patient characteristics, and outcomes for the final list of included studies (**Appendix A**). Following reconciliation between the two reviewers, a third reviewer will be included to reach consensus on any remaining discrepancies. For all outcomes of interest, information regarding point estimates, variability and uncertainty will be obtained. For PFS and OS, hazard ratios and associated information regarding uncertainty will be extracted. If results as reported as forest plot, the point estimate, and 95% confidence interval will be extracted using DigitizeIt® software version 2.1.4 (Bormisoft - Informer Technologies, Inc.). Kaplan Meyer curves will also be digitized 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 will be collected, if reported. Data collected may include 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 are reported will also be captured. Discontinuation due to adverse events may also be examined.

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

7.4.1.7. 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 will be restricted to trials published in English.

7.4.2. Network meta-analysis

7.4.2.1. Population, interventions, outcomes of interest

Relative treatment effects between the competing interventions for the treatment of EGFR+ NSCLC will be estimated by means of multiple NMAs, to be performed by line of treatment (i.e. 1L, 2L, and 2L+) conditional upon prior treatment (class) whenever feasible and relevant, based on data extracted from the studies identified with the SLR.(22,23) (For example, regarding 2L treatment, an NMA may be performed for a population that received a 1st generation TKI in the 1L, as well as an NMA for a population that received a 2nd generation TKI in the 1L setting.) The outcomes of interest are PFS, OS and adverse events.

7.4.2.2.   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 want to only consider the treatment effects of each trial. This consideration implies 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. If some interventions of interest are not part of the same network, then it is not possible to perform an indirect comparison of treatment effects of these interventions without a substantial risk of bias.(22-24)

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.(25,26)

In order to gauge the appropriateness of proceeding with an NMA, IVI's feasibility assessment will include: (i) an assessment of whether the RCT evidence for the interventions of interest forms 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 affect relative treatment effects across the direct comparisons of the evidence networks.

7.4.2.3.   Evaluation of consistency between direct and indirect comparisons

Prior to the actual NMA, the consistency between direct and indirect comparisons can be 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 will assess 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.(27)

7.4.2.4.   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 will be 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. If there is an imbalance in treatment potential effect modifiers across trials in the network, the NMA models may be extended by incorporating covariates (i.e. meta-regression analysis).(28)

7.4.2.5. Models, likelihood, priors

We will perform the NMA in a Bayesian framework. Analysis within the Bayesian framework involves data, a likelihood distribution, a model with parameters, and prior distributions. The model relates the data from the individual studies to basic parameters reflecting the (pooled) relative treatment effect of each intervention compared to an overall reference treatment. For each outcome and line of treatment of interest, fixed and random-effects models will be used. Because some heterogeneity is always anticipated, we will prefer a random-effects model. When few RCTs are available, however, heterogeneity estimation can become unreliable, and the use of conventional 'non-informative priors' in the Bayesian framework can result in artificially wide uncertainty intervals. Therefore, we may stabilize the heterogeneity estimation by using moderately informative heterogeneity variance priors or use fixed effects models instead.

7.4.2.5.1. *Progression-free survival and overall survival*

Traditional NMA for survival outcomes are based on hazard ratio (HR) estimates and rely on the proportional hazards assumption, which is implausible if the hazard functions of competing interventions cross.(29) Furthermore, separate meta-analyses of OS and of PFS data ignore the relationship between these outcomes. We will perform NMA of OS and PFS that is based on a tri-state (stable, progression, and death) transition model where time-varying hazard rates and relative treatment effects are modeled with known parametric survival functions or fractional polynomials according to **Equation 1** and illustrated in **Figure 3**.(29,30)

Figure 5. Relationship between stable disease (S), progression (P) and death (D) as used in the multistate network meta-analysis model
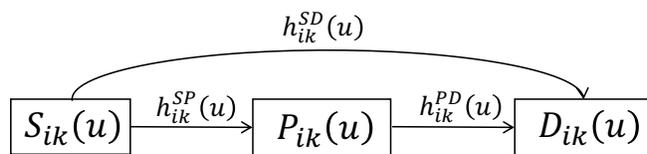
$S_{ik}(u)$ = progression -free (stable disease) in study i, treatment arm k at time u

$P_{ik}(u)$ = 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

$$ln\left(h_{ik}^{SP}(u)\right) = \begin{cases} \alpha_{0ik} + \alpha_{1ik}u^{p_1} + \alpha_{2ik}u^{p_2} & p_1 \neq p_2 \\ \alpha_{0ik} + \alpha_{1ik}u^p + \alpha_{2ik}u^p ln(u) & p_1 = p_2 = p \end{cases}$$

$$ln\left(h_{ik}^{SD}(u)\right) = \begin{cases} \alpha_{3ik} + \alpha_{4ik}u^{p_1} + \alpha_{5ik}u^{p_2} & p_1 \neq p_2 \\ \alpha_{3ik} + \alpha_{4ik}u^p + \alpha_{5ik}u^p ln(u) & p_1 = p_2 = p \end{cases} \qquad \text{eq. 1}$$

$$ln\left(h_{ik}^{PD}(u)\right) = \begin{cases} \alpha_{6ik} + \alpha_{7ik}u^{p_1} + \alpha_{8ik}u^{p_2} & p_1 \neq p_2 \\ \alpha_{6ik} + \alpha_{7ik}u^p + \alpha_{8ik}u^p ln(u) & p_1 = p_2 = p \end{cases}$$

$$\begin{pmatrix} \alpha_{0ik} \\ \alpha_{1ik} \\ \alpha_{2ik} \\ \alpha_{3ik} \\ \alpha_{4ik} \\ \alpha_{5ik} \\ \alpha_{6ik} \\ \alpha_{7ik} \\ \alpha_{8ik} \end{pmatrix} = \begin{pmatrix} \mu_{0i} \\ \mu_{1i} \\ \mu_{2i} \\ \mu_{3i} \\ \mu_{4i} \\ \mu_{5i} \\ \mu_{6i} \\ \mu_{7i} \\ \mu_{8i} \end{pmatrix} + \begin{pmatrix} \delta_{0,ik} \\ d_{1,1t_{ik}} - d_{1,1t_{i1}} \\ 0 \\ d_{3,1t_{ik}} - d_{3,1t_{i1}} \\ d_{4,1t_{ik}} - d_{4,1t_{i1}} \\ 0 \\ d_{6,1t_{ik}} - d_{6,1t_{i1}} \\ d_{7,1t_{ik}} - d_{7,1t_{i1}} \\ 0 \end{pmatrix}$$

$$\delta_{0,ik} \sim Normal\left(d_{0,1t_{ik}} - d_{0,1t_{i1}}, \sigma_0^2\right)$$

Where $p = \{-1,0,1\}$ and $t^0 = ln(u)$

$\alpha_{.ik}$ represent the scale and shape parameters of the log hazard functions in study i in treatment arm k. $\mu_{.i}$ reflect the study effects regarding the scale and shape parameters in each study i, $\delta_{0,ik}$ are the study specific true underlying relative treatment effects regarding the scale of the log hazard function for the transition from progression-free to progression, which are described by a normal distribution with an average effect for treatment t relative to reference treatment 1 ($d_{0,1t}$) and between study heterogeneity σ_0^2. The treatment effects regarding the scale of the log hazard for the transition from progression-free to dead ($d_{3,1t}$) and for the transition from progression to dead ($d_{6,1t}$) as well as the treatment effects regarding the first shape parameters of the functions describing the transition rates ($d_{1,1t}$, $d_{4,1t}$, and $d_{7,1t}$) are assumed to be fixed by treatment.

The following competing survival distributions will be considered using this multi-state NMA framework: Weibull ($p_1 = 0$), Gompertz ($p_1 = 1$), and 2nd order fractional polynomials ($p_1 = \{0,1\}$ and $p_1 = \{-1,0,1\}$). These 2nd order fractional polynomial models are extensions of the Weibull and Gompertz model and allow arc- and bathtub shaped hazard functions. To facilitate parameter estimation, we assume treatment to only have an impact on the time-varying transition rates from stable to progression and stable to death; we assume transition rates between progressive disease and death independent of treatment. For the relative treatment effects in the 2nd order fractional polynomial framework we will assume that treatment only has

an impact on two of the three parameters describing the hazard function over time (i.e. one scale and 1 shape parameter).

The model parameters will be estimated based on the number of patients in each of the three health states over time obtained from the published Kaplan-Meier (KM) curves for each arm of each trial included in the NMA. Accordingly, we will use a multinomial likelihood for the proportion of patients in each of the three health states at any point in time. These proportions 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.(30) The prior distributions for the model parameters are:

$$\begin{pmatrix} \mu_{0i} \\ \vdots \\ \mu_{8i} \end{pmatrix} \sim MVN\left(\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, T_\mu\right) \qquad T_\mu \sim MVN\begin{pmatrix} 10^4 & \cdots & 10^4 \\ \vdots & \ddots & \vdots \\ 10^4 & \cdots & 10^4 \end{pmatrix} \qquad \text{eq. 2}$$

$$\begin{pmatrix} d_{0,1t} \\ d_{1,1t} \\ d_{3,1t} \\ d_{4,1t} \\ d_{6,1t} \\ d_{7,1t} \end{pmatrix} \sim MVN\left(\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, T_d\right) \qquad T_d \sim MVN\begin{pmatrix} 10^4 & \cdots & 10^4 \\ \vdots & \ddots & \vdots \\ 10^4 & \cdots & 10^4 \end{pmatrix}$$

$$\sigma_0 \sim uniform(0,2)$$

7.4.2.5.2. *Adverse events*

For adverse events reported as the number of patients experiencing the event, the NMA will be performed on the proportion of patients experiencing the event of interest with a binomial likelihood and logit link.(23) For events reported as number of events per person-time by treatment (instead of number of persons with at least 1 event), we will use regression models with a log link function and Poisson likelihood. Normal non-informative prior distributions for the parameters were used with a mean of 0 and a variance of 10,000.

7.4.2.6.   **Model selection**

The deviance information criterion (DIC) will be used to compare the goodness-of-fit of competing evidence synthesis models.(31) DIC provides a measure of model fit that penalizes model complexity. In general, a more complex model will result 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 DIC of about 10 points can be considered meaningful. Regarding PFS and OS, the DIC estimates will be used to select a subset of appropriate evidence synthesis models which are available for the end-user to choose from when running the simulations. Furthermore, the DIC estimates will be used for model averaging (**See Section 7.3.2**). Regarding adverse events, the DIC will be used to inform whether estimates obtained with a fixed or random effects approach will be used in the model simulations.

7.4.2.7. Software

The parameters of the different models will be estimated using a Markov Chain Monte Carlo (MCMC) method implemented in the JAGS software package. JAGS will be run using R statistical software. (32)

7.4.3. *Meta-analysis of absolute effects*

For the synthesis of absolute treatment effects with the defined reference treatment by line of treatment, similar models will be used as for the NMA.

7.5. Model validation

Model validation will be performed using three primary steps. First, we will check the model for face validity by ensuring that simulated outcomes are consistent with current science and evidence. Second, we will perform unit tests to verify that the individual units of code that are used to simulate the model return expected results. Third, we will compare simulated results for PFS and OS according to different treatment sequences with the estimated PFS and OS curves obtained with the 1L evidence synthesis. In the future, depending on data available we may use both external validation and predictive validation to help refine the model. External validation can be performed by comparing outcomes simulated using our model to real-world outcomes and predictive validity will involve using our model to forecast future events and comparing our forecasted outcomes to the observed outcomes.

7.6. R Package

The model will be released as an R package on IVI's public GitHub repository, where all R and C++ code will be visible. An online model interface for an advanced user and a general audience value tool will also be released for interactive use. Documentation for the package will be available on a public website. The website will contain (i) a tutorial showing how to use the model, (ii) detailed technical documentation, and (iii) an application programing interface (API) reference describing all available functions, datasets, and classes.

Non-IVI modelers will be able use the R package for their own analyses by installing it from GitHub using standard commands available in the *devtools* package. Modelers can use the package as is or make direct changes to the code by forking (e.g. making a copy) of the repository.

Non-IVI modelers can choose to collaborate with IVI by suggesting that IVI incorporate the changes from their forked repository (e.g. make a "pull request"). IVI will assess the pull request and accept it if it improves the code (e.g. bug fixes, speed improvements) but does not impact model outcomes. Suggestions that impact model outcomes will be assessed with a formal process as described in **Section 7.7**.

7.7. Modified Delphi method to inform model revisions

Following the public release of the model, we will invite feedback and suggestions for improvements during a 2-month public comment period. Subsequently, a panel of subject matter experts will determine which of the received suggestions should be implemented by means of a modified Delphi process. The aim is to have the panel represent patients, patient advocates, oncologists, payers, and methodologist (e.g., economists, epidemiologists). Two rounds of surveys will be administered using a web-based survey tool. The focus of the initial survey will be on suggestions that require rather extensive resources to implement (such as elaborate programming or additional evidence collection) or are likely to have a substantial impact on the findings of a model analysis according to a review by the IVI team. Panel members will indicate on a four-point Likert scale ("strongly agree" to "strongly disagree") their level of agreement with each item on the questionnaire and provide reasons for their ratings, as well as provide a ranking of the suggestions that they believe are most important to incorporate. A "facilitator" will then provide an anonymized summary of the panel members' ratings and reasoning. After reviewing the anonymized summary, panel members will have an opportunity to update their answers with the second survey. The output of this process will be a list of prioritized model and interface modifications that will be incorporated into the next version based on the extent consensus has been achieved. New major versions of the model will be released every 12 months after completing the formal review process.

Appendix A: Systematic literature review

7.8. Eligibility criteria

7.8.1. Treatment effects

Table 3. 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 If the evidence-base regarding PFS and OS is too limited, studies reporting at least one of the following outcomes will be included in addition. • tumor response • DOR • EORTC QLQ-C30 • EORTC-LC13
Study design	RCTs
Other	English language

*Trials where the overall study population is an all-comer population, but subgroup results are reported for the target population of interest will be included; a trial or subgroup where the study population is a mixture of non-squamous and squamous patients will be included of >90% is non-squamous.

Table 4: 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 • nivolumab • pembrolizumab • atezolizumab • bevacizumab • platinum-based doublet therapy
Comparators	• placebo • 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 If the evidence-base regarding PFS and OS is too limited, studies reporting at least one of the following outcomes will be included in addition. • tumor response • DOR • EORTC QLQ-C30 • EORTC-LC13
Study design	RCTs
Other	English language

*Trials where the overall study population is an all-comer population, but subgroup results are reported for the target population of interest will be included; a trial or subgroup where the study population is a mixture of non-squamous and squamous patients will be included of >90% is non-squamous.

Table 5: 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 treatments*
Interventions	The following drugs as monotherapy or in combination with other drugs • nivolumab • pembrolizumab • atezolizumab • bevacizumab • platinum-based doublet therapy
Comparators	• placebo • 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 If the evidence-base regarding PFS and OS is too limited, studies reporting at least one of the following outcomes will be included in addition. • tumor response • DOR • EORTC QLQ-C30 • EORTC-LC13
Study design	RCTs
Other	English language

*Trials where the overall study population is an all-comer population, but subgroup results are reported for the target population of interest will be included; a trial or subgroup where the study population is a mixture of non-squamous and squamous patients will be included of >90% is non-squamous.

7.8.2. Utility

Table 6: PICOS criteria for review of utility estimates

PICOS	Criteria
Population	Adult patients with metastatic NSCLC
Interventions	No restriction on inclusion of studies based on interventions or comparators
Outcomes	- Utility measures (e.g. EQ-5d, HUI2, HUI-3, SF-6D) as a function of PFS, OS, TTP and adverse events; - mapping algorithms from a non-preference-based measure (generic or disease-specific measure) to a generic preference-based measure; - mapping algorithms between different generic preference-based health state utility values
Study design	- Reviews - Alternatively, primary studies
Other	English language

7.8.3. Cost-effectiveness modeling studies to obtain evidence on resource use, productivity, and cost

Table 7: PICOS criteria for studies providing information on resource use, productivity, and cost estimates

PICOS	Criteria
Population	Adult patients with metastatic NSCLC
Interventions	No restriction on inclusion of studies based on interventions or comparators
Outcomes	- Resource use - Cost - Productivity
Study design	US specific economic evaluations by means of integration of clinical evidence, resource use, and cost data using decision-analytic/mathematical modelling frameworks
Other	English language

7.9. Data extraction systematic review of treatment effects

7.9.1. Study characteristics

The following study characteristics will be extracted:

- Study name

- Study year

- Study author

- Sample size

- Study inclusion criteria

- Study exclusion criteria

- Location of study

- Number of study centers

- Masking status

- Study duration

- Follow-up period

- Treatment switching to intervention of interest after progression with control intervention

- Baseline stratification factors

- Diagnostic test used to confirm EGFR+ status

- Definition of outcomes

- Criteria and version number for response assessment

- Blinded versus non-blinded investigator assessment of response

7.9.2. Intervention characteristics

The following intervention characteristics will be extracted:

- Treatment regimen

- Treatment dose

- Method of administration

- Frequency of administration

- Duration of treatment

- Concomitant/background therapies

7.9.3. Patient characteristics

The following patient characteristics will be extracted:

- Age

- Gender

- Body Mass Index

- Race and ethnicity

- Smoking status and history

- ECOG performance score

- Current disease stage

- Tumor histology

- Vascular and local invasion

- Malignant pleural effusion

- Number of metastatic sites

- Brain metastasis status

- EGFR status

- KRAS status

- PD-L1 status

- Total mutation burden (TMB)

- Comorbidities

- Prior treatment experience

- Exon 19 deletion/L858R/other

- T790M resistance mutation

- Exon 20 insertion mutation

7.9.4.   Outcomes

The following outcomes will be extracted:

- OS

- PFS

- TTP

- Overall response rate (ORR)

- Complete response (CR)

- Partial response (PR)

- DOR

- Serious adverse events (Grade 3 and above)

- Treatment discontinuation

References

1. Jemal A BF, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA Cancer J Clin 2011:61:9-90.

2. D'Addario G, Fruh M, Reck M, Baumann P, Klepetko W, Felip E, et al. Metastatic non-small-cell lung cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2010;21 Suppl 5:v116-9.

3. Cetin K, Ettinger DS, Hei YJ, O'Malley CD. Survival by histologic subtype in stage IV nonsmall cell lung cancer based on data from the Surveillance, Epidemiology and End Results Program. Clin Epidemiol. 2011;3:139-48.

4. Lynch TJ, Bell DW, Sordella R, Gurubhagavatula S, Okimoto RA, Brannigan BW, et al. Activating Mutations in the Epidermal Growth Factor Receptor Underlying Responsiveness of Non–Small-Cell Lung Cancer to Gefitinib. N Engl J Med. 2004;350:2129-39.

5. Keeney, Ralph L, and Howard Raiffa. 1993. Decisions with Multiple Objectives: Preferences and Value Trade-Offs. Cambridge university press.

6. Thokala P, Devlin N, Marsh K, Baltussen R, Boysen M, Kalo Z, Longrenn T, Mussen F, Peacock S, Watkins J, Ijzerman M. Multiple Criteria Decision Analysis for Health Care Decision Making--An Introduction: Report 1 of the ISPOR MCDA Emerging Good Practices Task Force. Value Health. 2016;19:1-13.

7. Garrison LP Jr, Kamal-Bahl S, Towse A.Toward a Broader Concept of Value: Identifying and Defining Elements for an Expanded Cost-Effectiveness Analysis. Value Health. 2017;20:213-216.

8. Lakdawalla, D, Malani A, Reif J. The insurance value of medical innovation. J Public Economics 2017;145:94-102

9. Aalen OO, Johansen S. An empirical transition matrix for non-homogeneous Markov chains based on censored observations. Scandi J Statistics, 1978;5:141-150.

10. Putter H, Fiocco M, Geskus RB. Tutorial in biostatistics: competing risks and multi-state models. Stat Med 2007;26:2389-2430.

11. de Wreede LC, Fiocco M, Putter H. (2011). Mstate: an R package for the analysis of competing risks and multi-state models. J Stat Software 2011;38:1-30.

12. Jackson CH. Flexsurv: a platform for parametric survival modeling in R. J Stat Software. 2016;701-33.

13. Sanders GD, Neumann PJ, Basu A, Brock DW, Feeny D, Krahn M, Kuntz KM, Meltzer DO, Owens DK, Prosser LA, Salomon JA, Sculpher MJ, Trikalinos TA, Russell LB, Siegel JE, Ganiats TG. Recommendations for Conduct, Methodological Practices, and Reporting of Cost-effectiveness Analyses: Second Panel on Cost-Effectiveness in Health and Medicine. JAMA. 2016;316:1093-103.

14. Baio G, Dawid AP. Probabilistic sensitivity analysis in health economics. Stat Methods Med Res. 2015;24:615-34.

15. Claxton K, Sculpher M, McCabe C, Briggs A, Akehurst R, Buxton M, Brazier J, O'Hagan T. Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra. Health Econ. 2005;14:339-47

16. Black WC. The CE plane: a graphic representation of cost-effectiveness. Med Decis Making. 1990;10:212-4.

17. van Hout BA, Al MJ, Gordon GS, Rutten FF. Costs, effects and C/E-ratios alongside a clinical trial. Health Econ. 1994;3:309-19.

18. Briggs AH. A Bayesian approach to stochastic cost-effectiveness analysis. Health Econ. 1999;8:257-61.

19. Fenwick E, Claxton K, Sculpher M. Representing uncertainty: the role of cost-effectiveness acceptability curves. Health economics 2001;10:779–787.

20. Hoeting JA, Madigan D, Raftery AE, Volinsky CT. Bayesian model averaging: a tutorial. Statistical science. 1999;14:382-401.

21. Woods B, Sideris E, Palmer S, Latimer N, Soares M. NICE DSU Technical Support Document 19. Partitioned Survival Analysis for Decision Modelling in Health Care: A Critical Review. 2017 [Available from http://www.nicedsu.org.uk]

22. Jansen JP, Trikalinos T, Cappelleri JC, Daw J, Andes S, Eldessouki R, Salanti G. Indirect treatment comparison / network meta-analysis study questionnaire to assess relevance and credibility to inform healthcare decision-making: An ISPOR-AMCP-NPC Good Practice Task Force Report. Value Health. 2014;17:157-73.

23. Dias S, Ades AE, Welton NJ, Jansen JP, Sutton AJ. The Core Model. In: Network meta-analysis for decision making. Chichester UK: John Wiley & Sons Ltd, 2018; p.19-58

24. Dias S, Ades AE, Welton NJ, Jansen JP, Sutton AJ. Generalised Linear Models. In: Network meta-analysis for decision making. Chichester UK: John Wiley & Sons Ltd, 2018; p.93-153

25. Jansen JP, Schmid CH, Salanti G. Direct acyclic graphs can help understand bias in indirect and mixed treatment comparisons. J Clin Epidemiology 2012;65:798-807.

26. Jansen JP, Naci H. Is network meta-analysis as valid as standard pair-wise meta-analysis? It all depends on the distribution of effect-modifiers. BMC Medicine 2013;11:159.

27. Dias S, Ades AE, Welton NJ, Jansen JP, Sutton AJ. Checking for inconsistency. In: Network meta-analysis for decision making. Chichester UK: John Wiley & Sons Ltd, 2018; p.189-226

28. Dias S, Ades AE, Welton NJ, Jansen JP, Sutton AJ. Meta-regression for relative treatment effects. In: Network meta-analysis for decision making. Chichester UK: John Wiley & Sons Ltd, 2018; p.227-271.

29. Dias S, Ades AE, Welton NJ, Jansen JP, Sutton AJ. *Network meta-analysis of survival outcomes. In: Network meta-analysis for decision making. Chichester UK: John Wiley & Sons Ltd, 2018; p.293-322.

30. Jansen JP, Trikalinos T. Multivariate Network Meta-Analysis of Progression Free Survival and Overall Survival. Value Health. 2013;16:A617.

31. Dias S, Ades AE, Welton NJ, Jansen JP, Sutton AJ. Model fit, model comparison and outlier detection. In: Network meta-analysis for decision making. Chichester UK: John Wiley & Sons Ltd, 2018; p.59-91.

32. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2018 [URL http://www.R-project.org/]