Supplementary material to: The cost effectiveness of surveillance scanning strategies after curative treatment of Non-Small Cell Lung Cancer

Henri B. Wolff, Leonie Alberts, Elisabeth A. Kastelijn, Sherif Y. El Sharouni, Franz M. N. H. Schramel, Veerle M. H. Coupé
Table of Contents

Section 1: Microsimulation model
1.1 Simulation of Life Histories
1.2 Model functions
1.3 Calibration
1.4 Surveillance strategies
1.5 Costs per Item

Section 2: Sensitivity analyses
2.1 WHO discounting
2.2 Convergence test
2.3 Univariate Sensitivity analyses
2.4 Survival analyses

Section 3: All strategies (separate .xls file)
Section 1. Microsimulation model

1.1 Simulation of Individual Life histories of stage I NSCLC patients

The microsimulation model was built for the purpose of simulating individual disease histories of stage I NSCLC patients treated with curative intent. The model describes the growth and detection of metastases developing from the primary tumor. In addition, second primary lung tumors may develop. Metastases are assumed to start growing from a microscopic size that cannot be detected on a scan, and cause death when the combined tumor volume reaches the lethal threshold volume. The model simulates patients differing in the number of metastases, as seen in the literature, and in volume doubling time.¹⁻⁴

The model differentiates between young and healthy patients and older patients that are more likely to become untreatable,⁵ or to die of other causes.⁶ Treatability and life expectancy are the resultant of a complex combination of underlying characteristics which are not included in the model. However treatability and life expectancy are drawn from life tables and probability functions that reflect the heterogeneity of the Dutch patient population.

An overview of the model is given in Figure 1 of the main manuscript. However, to further illustrate how patients are simulated, some example life histories of hypothetical patients are given in Appendix Figure 1.
Appendix Figure 1  Example life histories. A) Patient A has no metastases after treatment of the primary tumor and does not develop a second primary tumor (SPLC), however, his health deteriorates, and it is decided to stop further scanning. This patient Dies of Other Causes (DOC) than cancer, or excess mortality. B) Patient has 1 oligo-recurrence from the primary tumor, which is symptomatically detected. Complications of the treatment causes death of the patient. C) Patient C has multiple metastases detected on a surveillance scan. Chemotherapy reduces growth, but does not cure the patient. The Patient Dies some time later Of Disease (DOD). D) Patient D develops SPLC, which is detected on a surveillance scan. This tumor is successfully removed, and the patient dies of other causes.
1.2 Model Functions

Part of the underlying tumor growth describes the process preceding detection of recurrences. As there is no data on this, some of the parameters that describe this underlying tumor growth have been calibrated which is further explained in section 1.3.

After curative treatment, patients are either disease free or have remaining undetectable metastatic disease. The latter occurs in a proportion $P_{\text{micrometa}}$ of simulated patients. To describe the growth of the metastases, an exponential tumor growth model using volume doubling time ($VDT$) in time $t$ (days$^{-1}$)$^7$ was chosen, because VDT is the most commonly used statistic in literature. Metastases from the same primary tumor are assumed to have the same volume doubling time, and their volumes can be pooled using geometric distribution rules. The total metastatic volume $V_M$ in a patient grows with a patient-specific VDT, which is randomly drawn from a quantile exponential distribution of VDTs based on pooling estimates from literature.$^8$

$$V_M(t) = V_M(0) \times 2^{t/VDT} . \ (1)$$

The model keeps track of the number of metastases in each patient. The total number of metastases ($M_{\text{total}}$) for each patient is drawn from a truncated normal distribution (such that metastatic patients have at least one metastasis). The relative sizes of the metastases follow a geometric distribution with size ratio $R$. $R$ is drawn from a beta distribution.

The total undetected metastatic tumor volume at baseline varies between patients. As a reference point, the time at which the largest metastasis reaches the ‘detectable size’ is drawn with a hazard rate $\lambda_{\text{detectable}}$. The detectable size is defined as a diameter of 5 mm assuming that small metastases have a spherical shape.

Once a metastasis passes the detectable size threshold ($t_{\text{det}}$), disease progression can be detected ($t_{\text{ass}}$) with a planned surveillance scan, or symptomatically, whichever occurs first. The time when a single metastasis becomes symptomatic is determined with

$$\lambda_{\text{symptom}} e^{-\lambda_{\text{symptom}} (t - t_{\text{det}})} , \ (2)$$

where $\lambda_{\text{symptom}}$ is the constant hazard rate, and $t_{\text{det}}$ is the time that metastasis $m$ has reached the detectable size and can become symptomatic. Therefore ($t - t_{\text{det}}$) describes the time that metastasis $m$ has been at risk of becoming symptomatic. For each patient, function 2 can be pooled for all metastases into a single cumulative hazard function that describes that one or more metastases in a single patient become symptomatic.

SPLC does not originate from the primary tumor and is therefore assumed to behave differently than the metastases of the primary tumor within the same patient. However, the VDT, $M$ and $R$ of SPLC and its possible metastases are randomly drawn from the same distributions as the metastases of the primary tumor.

Unlike metastases from the curatively treated primary tumor, SPLC could develop at any moment in time. Therefore, a hazard function is used with $\lambda_{\text{SPLC incidence}}$ to randomly draw a point in time when the patient would develop SPLC. The incidence rate is set around 1.5% per year based on literature.$^9$

SPLC can develop metastases with the same probability as the primary tumor, that is, $P_{\text{micrometa}}$. It was assumed that $P_{\text{micrometa}}$ is not dependent on the size of SPLC at time of detection, because most single-cell metastases already have disseminated from its tumor of origin time of diagnosis.$^{10}$ This would also explain findings that earlier detection of tumors did not result in fewer recurrences.$^{11}$ For each patient with SPLC, function 2 is also used to establish when a patient develops symptoms from the SPLC and its metastases.
After detection of a recurrence or SPLC, several options are considered. Curative treatment may be given to patients with a limited number (<3) of detected metastases (or SPLC) that have not reached the untreatable size determined with the probability $P_{\text{untreatable size}}$ (Table 1). These patients are also required to be healthy enough to undergo any kind of therapy.

The time when patients become untreatable was estimated from Dutch stage I-III NSCLC patients that were treated between 2001 and 2013. A clear age-dependent linear decline in surgery utilization was observed between ages 65 and 90, at which surgery utilization reached 0%. It was assumed that these patients were considered untreatable. In the model all patients are considered treatable at baseline. Therefore, a random time between 65 and 90, or the age of diagnosis of the primary tumor and 90 was used to estimate when patients become untreatable.

Curative treatment removes all tumors visible on a scan. Alternatively, healthier patients can get systemic therapy which reduces their VDT by multiplying VDT with $\beta_{\text{systemic}}$. The third option is best supportive care (BSC). In BSC, patients are not given any life-prolonging treatments, but instead treated to increase quality of life. Therefore, BSC is modeled as a separate state with its own costs and health utilities (Post recurrence).

In the model all remaining tumor volume keeps growing until the lethal volume ($\beta_{\text{lethal}}$) is reached. The function used is the same as the previous function used, with the exception that both the metastases of the primary tumor $M$ and SPLC add up to the total tumor volume used to calculate the time that $\beta_{\text{lethal}}$ is reached.6

$$\beta_{\text{lethal}} = V_M(t_{\text{DFS}}) \times 2^\left(t_{\text{DOD}}-t_{\text{DFS}}\right)/V_{\text{DT}}(M) + V_{\text{SPLC}}(t_{\text{DFS}}) \times 2^\left(t_{\text{DOD}}-t_{\text{DFS}}\right)/V_{\text{DT}}(\text{SPLC}). \ (3)$$

Here, $t_{\text{DFS}}$ is the reference time that the metastases of the primary tumor $M$ and/or SPLC are detected. This reference time and the volumes at that point in time $V_M$ and $V_{\text{SPLC}}$ are used, because from this point in time onwards, $\beta_{\text{systemic}}$ may alter the growth rates.

1.3 Calibration

All unobservable parameters have been estimated using calibration.8, 13 The mean squared error of 1000 simulations was used in combination with a univariate search algorithm, and repeated until the calibration target was matched up to 3 significant figure places.14, 15

The parameters were calibrated consecutively, in order of the influence on the model outcomes. First $\lambda_{\text{detectable}}$ was calibrated against 5 year DFS split in 6 month time intervals (Appendix figure 2). Secondly the total number $M$ of metastases per patient was calibrated against the proportion of patients with 3 or less metastases, which is assumed to be the proportion of patients curatively treated for oligo-metastases who are progression-free at 5 years.16-22 From all values for $M$ that fitted the calibration target, that value was selected that minimizes the maximum number of detected metastases (based on expert opinion). Thirdly the size ratio of consecutive metastases $R$ was calibrated against the ratio of proportions of curatively treated oligo-metastases with and without recurrences that are detected at a later point in time using a 5 year time horizon.19, 21, 22 Fourthly, the hazard of a single metastasis becoming symptomatic, $\lambda_{\text{symptom}}$, was calibrated against the ratio of symptomatically detected metastases to metastases detected by a scheduled scan.16, 17 The fifth parameter to be calibrated was $\beta_{\text{lethal}}$, the lethal tumor volume, with was calibrated to the OS of patients with recurrences that did not receive systemic therapy which was estimated from two RCTs.23, 24 Subsequently the effect of systemic therapy on tumor volume ($\beta_{\text{systemic}}$) was calibrated to OS of patients with systemic therapy.
Appendix Figure 2  Disease free survival of the patient populations and the calibrated model. Two Dutch datasets were pooled to obtain patient level data for VATS (green) and SBRT (blue). These patients were 1:1 propensity matched to minimize confounding. The resulting survival curves showed minimal differences (logrank 0.68). Subsequently, the VATS and SBRT patients were pooled and used for calibration of the underlying tumor growth model (red). Reprinted from Lung Cancer, 141, Henri B. Wolff, Leonie Alberts, Naomi van der Linden, Mathilda L. Bongers, Naomi E. Verstegen, Frank J. Lagerwaard, Carin A. Uyl-de Groot, Suresh Senan, Sherif Y. El Sharouni, Elisabeth A. Kastelijn, Franz M. N. H. Schramel, Veerle M. H. Coupé, Cost-effectiveness of stereotactic body radiation therapy versus video assisted thoracic surgery in medically operable stage I non-small cell lung cancer: A modeling study., 89-96, Copyright (2020), with permission from Elsevier.

1.4 Surveillance strategies

In this paper, 108 surveillance strategies were selected. The rationale for selecting these strategies was as follows. Firstly, in screening and surveillance many strategies are theoretically possible. If only a few are selected, the risk of choosing a non-optimal strategy may be high. Secondly, a surveillance strategy should be easy to implement. The early and late surveillance intervals as seen in Appendix Figure 3 are all derived from the variation as seen in the clinical practice and expert opinion, with the exception of the ‘None’ (no surveillance) strategy.
Appendix Figure 3 combinations of early and late phase surveillance intervals. Duplicate strategies such as when both early and late surveillance use the same intervals are removed, resulting in 108 unique strategies.

Duplicate strategies were removed. An example is 6 months early surveillance combined with 6 months late surveillance. Independent of the time when early surveillance switches to late surveillance, the time of the planned scans remain the same.

1.5 Costs per item

Appendix Table 1 Average costs for the categories curative therapy, systemic therapy, and best supportive care per cost item.28

|                      | curative therapy (N=108) | systemic therapy (N=112) | best supportive care (N=87) |
|----------------------|--------------------------|--------------------------|----------------------------|
| systemic therapy     | € 0                      | € 11,880                 | € 517                      |
| radio therapy (palliative) | € 0            | € 423                    | € 773                      |
| surgery              | € 6,421                  | € 0                      | € 0                        |
| concomitant therapy  | € 702                    | € 754                    | € 175                      |
| in-patient hospital days | € 6,168             | € 6,128                  | € 1,233                    |
| out-patient hospital visits | € 2,241        | € 3,076                  | € 281                      |
| intensive care       | € 1,151                  | € 278                    | € 0                        |
| day care             | € 175                    | € 837                    | € 23                       |
| imaging              | € 3,492                  | € 2,449                  | € 259                      |
| laboratory testing   | € 1,169                  | € 1,399                  | € 138                      |
| pathology            | € 220                    | € 55                     | € 2                        |
| telephone consultation | € 16                     | € 38                     | € 6                        |
| total average costs  | € 21,755                 | € 27,317                 | € 3,408                    |

Appendix Table 1 Average costs for patients with early stage (curative) and stage IV NSCLC. Best supportive care costs are estimated by calculating the average costs for patients that are ineligible for life-prolonging treatments. Systemic therapy patients are patients receiving chemotherapy or targeted therapy at baseline. As these patients are expected to receive best supportive care at some time after systemic therapy, the average best supportive care costs are subtracted from the average systemic therapy average costs to prevent counting the same costs twice. Summed average costs per category are used in the microsimulation model.
The costs shown in Appendix Table 1 were taken from a previous study and inflated to 2019. As this database contained limited data on systemic therapy usage after recurrent cancer, it was assumed that stage IV patients would incur the same costs (N=199). In the model costs for systemic treatment or best supportive care are charged at the start of treatment, and represents the costs of the entire time period during which such treatment is given.

Systemic therapy is a mixture of chemotherapy and targeted therapies. The following therapies were administered: carboplatin, cisplatin, docetaxel, etopocide, gemcitabine, paclitaxel, pemetrexed, bevacizumab, cetuximab, and erlotinib. The database consists of patients treated between 2009 and 2012, and these patients did not have access to immunotherapy at that time.

Section 2. Sensitivity Analyses

2.1 WHO discounting

All analyses in the main manuscript have been repeated with WHO discount rates (3% for costs and 3% for effects). In contrast, the Dutch discount rates are 1.5% for costs and 4% for effects. As a result, WHO discounting leads to higher costs and lower QALYs for all strategies. Furthermore, WHO discount rates cause all cost-effective strategies that form the cost-effectiveness frontier to have higher ICERS.

WHO discounting affects which strategies are on the cost-effectiveness frontier. Strategies EARLY_0 SWITCH_5 LATE_2, EARLY_0 SWITCH_3 LATE_2 and EARLY_0 SWITCH_2 LATE_1 are no longer on the cost-effectiveness frontier when WHO discount rates are used, and strategy EARLY_0 SWITCH_1 LATE_1.5 instead lies on the frontier when using WHO discount rates (see Appendix Table 2, and Table 2 in the main manuscript).

The strategies that were identified in our base-case analysis remain on the cost effectiveness frontier when using WHO discount rates. These are: EARLY_0 SWITCH_2 LATE_2, which was identified as the best alternative to no surveillance at the willingness to pay threshold of €50,000/QALY, and EARLY_0 SWITCH_1 LATE_2 which is the best option at the willingness to pay threshold of €80,000/QALY when using Dutch discount rates.

Although some strategies did switch to the frontier or away from the frontier, the order of both costs and effectiveness of strategies remain unaffected with WHO discounting as can be seen in the cost-effectiveness frontier (Appendix Figure 4 and Figure 2 in the main manuscript).

Further analyses were performed to identify the willingness to pay (WTP) regions where expected loss curves were less than € 100 from the ELC-frontier. Even with this conservative approach, 7 new strategies were identified that were a minimal distance away from the cost-effectiveness frontier, and in some WTP regions up to nine strategies were simultaneously within the € 100 region (Appendix Table 2).
| Number | Early surveillance frequency | Switch after | Late surveillance frequency | Costs (€ pp) | QALYs pp | ICERS (WTP/1000) | Close to frontier in this WTP range | Costs (€ pp) | QALYs pp | ICERS (WTP/1000) | Close to frontier in this WTP range |
|--------|-----------------------------|--------------|-----------------------------|--------------|---------|--------------------|-----------------------------------|--------------|---------|--------------------|-----------------------------------|
| 1      | None                        | -            | None                        | 9,892        | 5.721   | *                  | 0 – 55                            | 10,107       | 5.201   | *                  | 0 – 71                             |
| 2      | None                        | 5 years      | 2 years                     | 10,946       | 5.742   | 50,951            | 46 – 63                           | 11,268       | 5.218   | -                  | 65 – 71                            |
| 3      | None                        | 3 years      | 2 years                     | 11,110       | 5.745   | -                  | 48 – 66                           | 11,437       | 5.221   | -                  | 64 – 75                            |
| 4      | None                        | 2 years      | 2 years                     | 11,184       | 5.745   | -                  | 50 – 66                           | 11,523       | 5.221   | -                  | -                                 |
| 5      | None                        | 3 years      | 1½ years                    | 11,306       | 5.748   | 53,103            | 48 – 76                           | 11,638       | 5.224   | -                  | 64 – 86                            |
| 6      | None                        | 4 years      | 1½ years                    | 11,395       | 5.749   | -                  | 50 – 65                           | 11,740       | 5.224   | -                  | -                                 |
| 7      | None                        | 2 years      | 2 years                     | 11,549       | 5.753   | 57,293            | 43 – 91                           | 11,885       | 5.228   | -                  | 67,136                            |
| 8      | None                        | 3 years      | 1½ years                    | 11,655       | 5.753   | 54 – 76            | 12,003                            | 5.228        | -       | -                  | 81 – 117                           |
| 9      | None                        | 1 year       | 2 years                     | 11,860       | 5.757   | 71,639            | 53 – 101                          | 12,197       | 5.232   | 83,422             | 66 – 130                           |
| 10     | None                        | 1 year       | 1½ years                    | 12,410       | 5.763   | -                  | 74 – 105                          | 12,759       | 5.237   | 110,728            | 91 – 137                           |
| 11     | None                        | 2 years      | 1½ years                    | 11,980       | 5.758   | -                  | 62 – 100                          | 12,332       | 5.232   | -                  | 81 – 117                           |
| 12     | None                        | 3 years      | 1 year                      | 12,166       | 5.760   | 71 – 100           | 12,556                            | 5.234        | -       | -                  | -                                 |
| 13     | None                        | 1 year       | 1½ years                    | 12,410       | 5.763   | -                  | 74 – 105                          | 12,759       | 5.237   | 110,728            | 91 – 137                           |
| 14     | None                        | 2 years      | 1 year                      | 12,623       | 5.766   | 89,806            | 78 – 107                          | 13,015       | 5.239   | -                  | 105 – 136                          |
| 15     | None                        | 2 years      | ½ year                      | 14,225       | 5.781   | 100,779           | 95 – 145                          | 14,693       | 5.252   | 130,479            | 124 – 165                          |
| 16     | None                        | 6 months     | 2 years                     | 15,046       | 5.777   | -                  | -                                 | 15,346       | 5.249   | -                  | -                                 |
| 17     | None                        | 1 year       | ½ year                      | 15,292       | 5.789   | 132,764           | 120 – 176                         | 15,750       | 5.259   | 150,884            | 137 – 204                          |
| 18     | None                        | 3 months     | 5 years                     | 19,259       | 5.813   | 171,394           | 167 – inf                         | 19,702       | 5.278   | 202,345            | 197 – inf                          |

**Appendix Table 2** Dutch and WHO discounted costs (1.5% and 3% discount rates) and QALYs (4% and 3% discount rate) per person, as well as ICERs for strategies on the cost-effectiveness frontier, both when using Dutch and WHO discount rates. Additionally, the range of Willingness to Pay (WTP/1000) values where strategies are close to the ELC frontier (< € 100 away) are also reported. Strategies on the frontier are colored grey. Strategy 10 is not on the frontier when using Dutch discount rates, while strategies 2, 3 and 6 are no longer on the frontier when WHO discount rates are used. Strategies 11-16 are not on the frontier with either Dutch or WHO discount rates, but are identified as close to the WTP frontier. None of the cost-effective strategies has early surveillance except for strategy 9. This effectively means that the other strategies start scanning at the switch point. A complete overview of all 108 strategies is shown in Appendix section 3.

Theoretically, the strategy with the lowest expected loss is defined as the optimal risk-neutral strategy, and the strategies that are at a distance from the frontier (even if this distance is small), are by definition risk seeking. However, the risk of €100 loss distance from the ELC frontier is extremely small, and these strategies could be considered equivalent to those on the frontier.

These additional strategies are included in appendix table 2, and are very similar to the strategies that were previously selected; for instance strategy EARLY_0_SWITCH_4_LATE_2, which lies between EARLY_0_SWITCH_5_LATE_2, and EARLY_0_SWITCH_3_LATE_2. Combined they show a pattern of declining switch times (from 5 to 1 years) and declining late surveillance frequencies (from 2 to ½ years).
Appendix Figure 4  Cost effectiveness plane and frontier using WHO discount rates (3%). The cost-effectiveness frontier (line) connects all potentially cost-effective strategies that dominate all other strategies directly or via extended dominance (squares). All other strategies are either dominated (circles), meaning that they are more expensive and less effective than a strategy on the frontier, or subject to extended dominance (triangles), meaning that a combination of strategies on the frontier can be found that leads to higher effectiveness at the same costs, or lower costs at equal effectiveness. The numbers of the strategies on the frontier correspond to the numbering of the strategies shown in Appendix Table 2.

Appendix Figure 5 depicts the Cost Effectiveness Acceptability Curve (CEAC), and the Expected Loss Curve (ELC) for WHO discount rates.
Appendix Figure 5 Probabilistic Sensitivity Analysis (PSA) outcomes when using WHO discount rates (3%). A) The Cost Effectiveness Acceptability Curve shows the proportion of simulations in which each of the strategies has the maximum Net Monetary Benefit (NMB) as a function of the willingness to pay threshold. B) The Expected Loss Curve shows the expected difference between the NMB of a strategy and the maximum achieved NMB within the same PSA parameter set as a function to the willingness to pay threshold. This results in the expected value of the losses if choosing certain strategy. The dashed black lines connect the strategies on the cost-effectiveness frontier.
2.2 Convergence test

The PSA consists of 1000 simulations of 108 strategies, each consisting of 100,000 patients. Expected values of the 1000 simulations are used as the outcome of the model. The choice to use 1000 parameter-sets is common, but arbitrary. As 108 strategies are compared, it is good practice to test how stable the outcomes of the model are to the number of simulations.

For this purpose, a convergence plot of the NMB of strategies on the cost-effectiveness frontier was made, inspired by the paper of Hatswell et al. For this purpose the 1000 PSA simulations used in this paper were randomized in their order. Subsequently, the average net monetary benefit at a willingness to pay of 50,000 €/QALY was calculated over the accumulating number of PSA simulations. These increasing numbers of simulations were plotted against average net monetary benefit for each strategy on the frontier see Appendix Figure 6.

Appendix Figure 6  Convergence plot. Average net monetary benefit per strategy at willingness to pay of 50,000 €/QALY versus increasing numbers of randomly drawn simulations.

Appendix figure 6 shows that the Net Monetary Benefit becomes relatively stable around 100 simulations, and this especially holds for the ordering of the strategies and the difference between the strategies. As a result, the expected losses will be unaffected. It is therefore very unlikely that increasing the numbers of simulations in the PSA would make a difference in the outcomes of our study.

2.3 Univariate sensitivity analyses

The PSA estimates the effect of the combined parameter uncertainty on the outcomes of the model, but does not differentiate how individual parameters affect these outcomes. In this particular microsimulation study, all strategies share the same model parameters. Therefore, some model parameters may affect all strategies equally, and will not affect the difference between strategies, or their incremental cost-effectiveness.

The reason behind the following univariate sensitivity analyses is to identify model parameters that can have an effect on incremental cost-effectiveness, and change the choice of optimal strategy. As all strategies share the same parameters, but do have increasing numbers of scans, the difference in resource use (meaning application of scans and treatments) was analyzed: minimum resource use
(EARLY_0_SWITCH_LATE_0) versus maximum resource use (strategy EARLY_3_SWITCH_5_LATE_0.5, which is maximum scanning). If the incremental cost-effectiveness ratio for maximum scanning versus no scanning does not change, a parameter may still affect both costs or effects of minimum and maximum scanning equally, but it is unlikely that it does affect the decision for a particular surveillance strategy.

As the model was analyzed only through a probabilistic simulation (PSA), a standard deterministic sensitivity analysis where only one parameter is varied compared to a deterministic base-case analysis is not appropriate. Instead, the effect of single parameters on the PSA outcomes was investigated by using linear regression of the incremental QALYs and incremental costs as the outcome variables and using each parameter shown in Table 1 separately as the predictors.

Only significant predictors ($p < 0.05$) were selected and the 2.5 and 97.5 percentile intervals of the predicted outcomes are shown in a tornado plot (Appendix Figure 7). The parameters in the tornado plots are ordered according to their effect on incremental differences between no- and maximum-surveillance strategy.

**Appendix Figure 7** Tornado plot of the impact of changes in model parameters on the difference in outcomes between no scanning and the most intensive scanning strategy. The effect of model parameters (shown in Table 1 in the main manuscript) on the incremental QALYs (A) and costs (B) was estimated using linear regression. Insignificant predictors ($p > 0.05$) are not shown. The bars represent the 95% predicted range of differences in QALYs and costs. QoL stands for Quality of life, which represent the health utilities for each specific health state.
This univariate sensitivity analysis shows that the difference between the two most extreme strategies is most sensitive to varying the SPLC incidence rate ($\lambda_{\text{SPLC incidence}}$), and the costs of CT scans. A high SPLC incidence rate would make increased surveillance more expensive and more effective. SPLC incidence and a low average total number of metastases are linked to an increased number of patients that may benefit from curative therapy for SPCL or oligo-metastases, which could explain why late surveillance is more cost-effective than early surveillance. The other identified parameters affect survival.

All tornado plots show positive values on the x-axis, meaning that for every single parameter change, extensive surveillance remains more expensive and more effective than no surveillance. It should be noted that this reflects the impact on the difference between the two most extreme strategies (no surveillance, maximum surveillance). For strategies with smaller differences in surveillance intensity, the differences in costs and effects will also be smaller.

2.4 Survival analyses

The cost-effectiveness plane gives insight in the total QALYs per person when a patient cohort is followed over the entire lifetime. However, it does not show during what time the patients benefit most from surveillance. To give additional insight into how surveillance strategies affect survival, 5 year DFS and OS were visualized using a Kaplan Meier curve for the most intensive strategy of ‘surveillance scanning once per 3 months’, the intermediate strategy ‘surveillance scanning once per year’, and the ‘no surveillance scans’ strategy.

Analyses with more strategies included (data not shown) show that DFS of all surveillance strategies fall between the DFS curves of no scanning strategy (symptomatic detection) and the maximum scanning strategy (every 3 months).

Appendix figure 8A suggests that the maximum difference in the time of detection of a recurrence is quite limited. This difference is the largest in the first two years after curative therapy of the primary tumor, which is in line with the findings in literature.

The difference in overall survival between the surveillance strategies is much smaller, despite the fact that the model included the potential advantages of early detection. That is: a longer period of reduced growth caused by systemic therapy, and increased risks of becoming untreatable at later detection and increased risk of excess mortality due to treatment.

In the case of systemic therapy, the difference in detection time would lead to an additional period in which the growth of metastases is reduced due to the treatment. With these assumptions it will be very unlikely that this would lead to an additional survival advantage that is larger than the average survival advantage that is gained in the current clinical practice. This strengthens the argument that some of the conflicting results reported in the literature may have been caused by lead time bias.
Appendix Figure 8 A) Disease Free Survival and B) Overall Survival of the strategies “no surveillance” (blue), “1 scan per 3 months” (green), and “1 scan per year” (red). The plot is limited to 5 year survival, although patients in the model may live much longer.
References

1. Frelinghuysen M, Fest J, Van der Voort Van Zyp NC, et al. Consequences of Referral Time and Volume Doubling Time in Inoperable Patients With Early Stage Lung Cancer. *Clin Lung Cancer* 2017; 18: e403-e409. DOI: 10.1016/j.cllc.2017.05.002.

2. Honda O, Johkoh T, Sekiguchi J, et al. Doubling time of lung cancer determined using three-dimensional volumetric software: comparison of squamous cell carcinoma and adenocarcinoma. *Lung Cancer* 2009; 66: 211-217. DOI: 10.1016/j.lungcan.2009.01.018.

3. Mackintosh JA, Marshall HM, Yang IA, et al. A retrospective study of volume doubling time in surgically resected non-small cell lung cancer. *Respirology* 2014; 19: 755-762. DOI: 10.1111 resp.12311.

4. Murai T, Shibamoto Y, Baba F, et al. Progression of non-small-cell lung cancer during the interval before stereotactic body radiotherapy. *Int J Radiat Oncol Biol Phys* 2012; 82: 463-467. DOI: 10.1016/j.ijrobp.2010.10.001.

5. Dutch Cancer Registry managed by the Comprehensive Cancer Centre the Netherlands. 2015.

6. Wolff HB, Alberts L, van der Linden N, et al. Cost-effectiveness of stereotactic body radiation therapy versus video assisted thoracic surgery in medically operable stage I non-small cell lung cancer: A modeling study. *Lung Cancer* 2020; 141: 89-96. DOI: 10.1016/j.lungcan.2020.01.011.

7. Schwartz M. A Biomathematical Approach to Clinical Tumor Growth. *Cancer* 1961; 14: 1272-1294. DOI: Doi 10.1002/1097-0142(196111/12)14:6<1272::Aid-Cncr2820140618>3.0.Co;2-H.

8. Wolff HB, Alberts L, Kastelijn EA, et al. Prediction of microscopic metastases in patients with metachronous oligo-metastases after curative treatment of Non-Small Cell Lung Cancer. *bioRxiv* 2019: 693747. DOI: 10.1101/693747.

9. Chang CF and Gould M. Playing the odds: lung cancer surveillance after curative surgery. *Curr Opin Pulm Med* 2017; 23: 298-304. 2017/04/14. DOI: 10.1097/MCP.0000000000000381.

10. Klein CA. Parallel progression of primary tumours and metastases. *Nat Rev Cancer* 2009; 9: 302-312. DOI: 10.1038/nrc2627.

11. Bach PB, Jett JR, Pastorino U, et al. Computed tomography screening and lung cancer outcomes. *JAMA* 2007; 297: 953-961. 2007/03/08. DOI: 10.1001/jama.297.9.953.

12. Mollberg NM and Ferguson MK. Postoperative surveillance for non-small cell lung cancer resected with curative intent: developing a patient-centered approach. *Ann Thorac Surg* 2013; 95: 1112-1121. 2013/01/29. DOI: 10.1016/j.athoracsur.2012.09.075.

13. Wolff HB, Alberts L, Kastelijn EA, et al. Differences in Longitudinal Health Utility between Stereotactic Body Radiation Therapy and Surgery in Stage I Non-Small Cell Lung Cancer. *J Thorac Oncol* 2018; 13: 689-698. DOI: 10.1016/j.jtho.2018.01.021.

14. Rutter CM, Zaslavsky AM and Feuer EJ. Dynamic Microsimulation Models for Health Outcomes: A Review. *Medical Decision Making* 2011; 31: 10-18. DOI: 10.1177/0272989x10369005.

15. Stout NK, Knudsen AB, Kong CY, et al. Calibration Methods Used in Cancer Simulation Models and Suggested Reporting Guidelines. *Pharmacoeconomics* 2009; 27: 533-545. DOI: Doi 10.2165/11314830-00000000-00000.

16. Lou F, Huang J, Sima CS, et al. Patterns of recurrence and second primary lung cancer in early-stage lung cancer survivors followed with routine computed tomography surveillance. *J Thorac Cardiovasc Surg* 2013; 145: 75-81. DOI: 10.1016/j.jtcvs.2012.09.030.

17. Westeel V, Choma D, Clement F, et al. Relevance of an intensive postoperative follow-up after surgery for non-small cell lung cancer. *Ann Thorac Surg* 2000; 70: 1185-1190.

18. Congedo MT, Cesario A, Lococo F, et al. Surgery for oligometastatic non-small cell lung cancer: long-term results from a single center experience. *J Thorac Cardiovasc Surg* 2012; 144: 444-452. DOI: 10.1016/j.jtcvs.2012.05.051.
19. Hishida T, Yoshida J, Aokage K, et al. Postoperative oligo-recurrence of non-small-cell lung cancer: clinical features and survival. *Eur J Cardiothorac Surg* 2016; 49: 847-853. DOI: 10.1093/ejcts/ezv249.

20. Kwint M, Walraven I, Burgers S, et al. Outcome of radical local treatment of non-small cell lung cancer patients with synchronous oligometastases. *Lung Cancer* 2017; 112: 134-139. DOI: 10.1016/j.lungcan.2017.08.006.

21. Shimada Y, Saji H, Kakihana M, et al. Survival outcomes for oligometastasis in resected non-small cell lung cancer. *Asian Cardiovasc Thorac Ann* 2015; 23: 937-944. DOI: 10.1177/0218492315596463.

22. Torok JA, Gu L, Tandberg DJ, et al. Patterns of Distant Metastases After Surgical Management of Non-Small-cell Lung Cancer. *Clin Lung Cancer* 2017; 18: e57-e70. DOI: 10.1016/j.clc.2016.06.011.

23. Non-Small Cell Lung Cancer Collaborative G. Chemotherapy and supportive care versus supportive care alone for advanced non-small cell lung cancer. *Cochrane Database Syst Rev* 2010: CD007309. DOI: 10.1002/14651858.CD007309.pub2.

24. Zhong C, Liu H, Jiang L, et al. Chemotherapy plus best supportive care versus best supportive care in patients with non-small cell lung cancer: a meta-analysis of randomized controlled trials. *PLoS One* 2013; 8: e58466. DOI: 10.1371/journal.pone.0058466.

25. O'Mahony JF, Naber SK, Normand C, et al. Beware of Kinked Frontiers: A Systematic Review of the Choice of Comparator Strategies in Cost-Effectiveness Analyses of Human Papillomavirus Testing in Cervical Screening. *Value Health* 2015; 18: 1138-1151. 2015/12/22. DOI: 10.1016/j.jval.2015.09.2939.

26. Erb CT, Su KW, Soulos PR, et al. Surveillance Practice Patterns after Curative Intent Therapy for Stage I Non-Small-Cell Lung Cancer in the Medicare Population. *Lung Cancer* 2016; 99: 200-207. 2016/08/28. DOI: 10.1016/j.lungcan.2016.07.017.

27. Sharma G, Nishi SP, Lin YL, et al. Pattern of Imaging after Lung Cancer Resection. 1992-2005. *Ann Am Thorac Soc* 2016; 13: 1559-1567. 2016/06/01. DOI: 10.1513/AnnalsATS.201511-768OC.

28. van der Linden N, Bongers ML, Coupe VM, et al. Costs of non-small cell lung cancer in the Netherlands. *Lung Cancer* 2016; 91: 79-88. DOI: 10.1016/j.lungcan.2015.10.015.

29. Hatswell AJ, Bullement A, Briggs A, et al. Probabilistic Sensitivity Analysis in Cost-Effectiveness Models: Determining Model Convergence in Cohort Models. *Pharmacoeconomics* 2018; 36: 1421-1426. 2018/07/28. DOI: 10.1007/s40273-018-0697-3.