Cost-effectiveness of lung cancer screening programs: comparing the NLST and NELSON screening

Authors: Gertjan Mantjes, Lotte M.G. Steuten[1], Maarten J. Ijzerman[2], Harry J.M. Groen[3], Joshua A. Roth[1], Bernardo H.L. Goulart[1], Richard D. Kim[1], Sabine Siesling[2,4]

[1] Hutchinson Institute For Cancer Outcomes Research, Fred Hutchinson Cancer Research Center, Seattle, United States
[2] Department Health Technology and Services Research, University of Twente, Enschede, the Netherlands
[3] Department of Pulmonary Diseases, University Medical Center Groningen, Groningen, the Netherlands
[4] Department research, Netherlands Comprehensive Cancer Centre, Utrecht, the Netherlands

Introduction

The United States Preventive Service Task Force recommended lung cancer screening in 2014 in the United States (US) based on outcomes of the National Lung Cancer Screening Trial (NLST).[1] The NLST showed a 20% reduction in lung cancer deaths. In February 2015, Medicare announced to reimburse lung cancer screening for high risk enrolllees.[4] However, positive predictive value of screening was 4.4%, whereas 96.4% of all screening results had a false-positive result.[2,3]

The goal of a screening program is to detect tumors in a stage when patients are asymptomatic, treatment is most successful and treatment costs are lower.[5] Screening for lung cancer is feasible for three different reasons.[6,7] First, lung cancer has a poor prognosis with an average survival rate of 17%.[8] A significant increase in survival rate can be achieved in case a tumor is detected in early stage. Stage I has an average five-year survival rate of 65.9% versus 0.1% of a stage IV tumor.[9,10] Currently, only 15% of tumors are detected in stage I to III.[1] A second opportunity for screening is the easy identified risk group. Tobacco smoking is the most important cause of lung cancer. The probability for heavy smokers to develop lung cancer is 33% higher when compared to non-smokers.[11–13] The American Cancer Society reports a life time probability of developing lung cancer of 6.3% for women and 7.7% for men. Smoking accounts for 87% of lung cancer deaths among men and 70% of lung cancer deaths among women.[14] The third reason for introducing screening is the decrease in short-term costs in case tumors can be detected in an earlier stage. Healthcare costs attributable to lung cancer amount to 8.4% ($449.5 million) and 9.7% ($12.1 billion) of total cancer costs in the Netherlands and the US respectively.[15,16] This is mostly made up out of treatment costs and depends strongly on tumor stage. In 2000, treatment costs of a 72 year old patient in the US in stage IV were 23.8% higher than treatment costs in stage I/II.[7]

Several other countries have started lung cancer screening trials.[3,17–22] The NEderlands-Leuvens Longkanker ScreeningsONderzoek' (NELSON) is carried out in the Netherlands and Belgium.[3,17] NELSON started in 2003, and like the NLST, used low-dose computed tomography (LDCT) as a primary screening diagnostic. Lung cancer screening is not yet implemented in the Netherlands and Belgium. In contrast to

the NLST screening program, the NELSON has shown to yield a positive predictive screen value of 40.4% and only 1.2% of all screening results were false-positive.[17]

It is unknown which elements cause the difference in detection rates between the strategies. Neither will costs be comparable across screening strategies. Hence, cost-effectiveness will be the primary outcome in this study in order to assess a full view of the effects of a screening program. First, the influence of the screening elements on the expected cost-effectiveness of the screening strategies as studied in the NLST and NELSON trials are determined. Second, the drivers of cost-effectiveness in the NLST and NELSON screening strategies are compared. Third, it is explored to what extent the trial-based estimates of cost-effectiveness would hold in a real-world setting, by populating the model with a sample of lung cancer screening registry data. These analyses will be performed to create an advice to improve cost-effectiveness of screening strategies in both settings.

Methods

Overview

A decision tree is built to assess the cost-effectiveness of screening. To assess which part of the screening programs has the biggest influence on the cost-effectiveness, the model is divided in three modules: (1) the population module, which contains all parameters that specify the characteristics of the eligible and responsive population based on study protocols by both trials, (2) the screening algorithm module, which contains all parameters pertaining to the measurement and categorization of nodules, and (3) the diagnostic work-up module, which includes parameters describing the utilization and tariffs of diagnostics after a positive screening result. Data from a lung cancer screening registry in Seattle (WA) are used to mirror the trial-based cost-effectiveness to a real-world setting.

Population and setting

The base case population are participants from the NLST and NELSON, which are a number of 26,722 and 7,915 respectively.[17,23] At the start of the model, all participants are asymptomatic, but at high risk of developing lung cancer. Participants of the control arm of both trials are not included, because information about the control arm in NELSON is not published yet.[17] The NLST was carried out in the US health care system; the trial data from the NELSON is only from the Dutch health care system, because the Belgian data is not public yet.[17] Population characteristics can differ due to inclusion and exclusion criteria of screening programs (Appendix - Table 1).[17,24] Beside differences in inclusion and exclusion criteria, participants characteristics can differ due to recruitment methods. The NLST did not use any methods to prevent a healthy-volunteer bias, in contrast to NELSON.[3,17] The real-world population is like the trial population subject to the response of the population, but can also be influenced due to insurance rates.[25–29] Effects of insurance rates cannot be seen in the characteristics of trial participants.

The Fred Hutchinson Lung Cancer Data Registry (Seattle, WA) was founded to collect data on the use and performance of lung cancer screening with LDCT.[30] The data registry is used to collect data from patients in real-world practice.[31] A total of 114 participants were enrolled in the data registry, of which 56 participants were screened at the time of analysis (Appendix - Table 2).

Study perspective and interventions

Third party payers made the decision whether or not to implement a screening program in the US, justifying the payer perspective. This perspective is used in both settings to keep the results comparable. The program costs are based on costs of LDCT scans and diagnostic work-ups.

The screening programs of the NLST and NELSON were aimed on finding tumors in early stage. Both strategies used LDCT as the primary screening method. The NLST used manual measurement to determine the diameter of the nodules.[24] The NELSON used semi-automatic volume measurement software to calculate the diameter, volume and the volume doubling time.[17] The volume doubling time could only be calculated if a previous scan was performed.

Both strategies used thresholds to categorize the screening results (Appendix - Table 3). The NLST considered two categories: 'positive' and 'negative'. NELSON additionally included the 'indeterminate' category. When a screening result was positive, a diagnostic work-up followed. When a screening result was negative, the participant was invited to the next screening round. If a screening result was indeterminate, the participant was invited to a follow-up scan, before the next screening round. In this follow-up scan, only positive or negative screening results were possible.

Time horizon and discounting

This study analyzed the first three screening rounds of both trials. Patient follow-up in the NLST finished after three rounds; NELSON included a fourth round from which data is not analyzed yet. NLST used a one year duration for every screening round, resulting in a total trial duration of three years. NELSON increased the time interval after every round, leading to a total time span of 5.5 years (Figure 1).

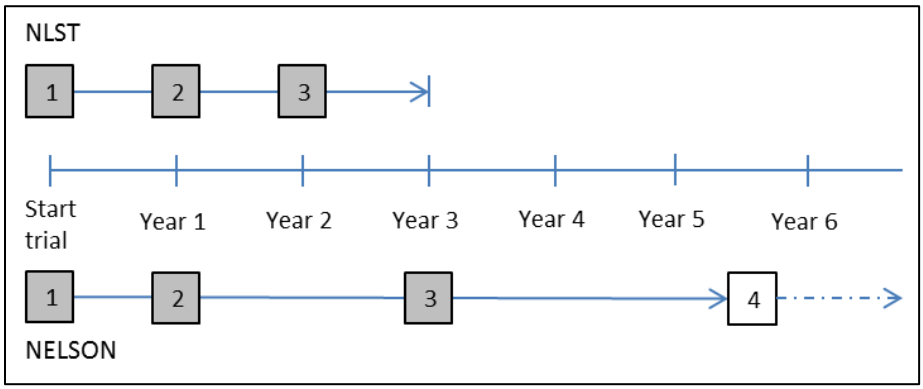


Figure 1: Screening rounds duration base case

The discount rates are based on the country in which the trials are set. Using these setting-based rates gives the opportunity to assess the effect of the elements per screening strategy per setting. The NLST is discounted for 3% a year on costs and health outcomes.[32] The NELSON is discounted for 4% a year for costs, and 1.5% a year for health outcomes.[33]

Health outcomes

The quality-adjusted life year (QALY) is the primary health outcome of the model. QALYs are calculated by weighing the length of life by the health-related quality of life. The quality of life is expressed in utilities, which are based on the screening results from participants.

Life-years are determined by calculating the number of participants entering the trial and the time span of the trial. This is adjusted for the chance of dying due to lung cancer and the chance of dying of another cause. The chance of dying due to lung cancer is calculated using the chance of developing lung cancer combined with the weighted average of survival rates to compare for the different tumor stages (Appendix - Table 4).[10,17,34,35] The probability of dying due to other causes are derived from the Centers for Disease Control and Prevention (CDC) for NLST and Statistics Netherlands (CBS) for NELSON (Appendix - Table 5).[36,37]

Utilities used are estimated for a period longer than 12 months by Black et al. (Appendix - Table 4).[38] Black used the Short Form Health Survey SF-36 to derive the utilities.[39] No adjustments are made for false-positive results, because there is still uncertainty about the influence.[40]

Resources and costs

To determine the total trial costs, the unit cost and utilization of screening and diagnostic work-up need to be gathered.

The utilization of LDCT scans is derived from the study results of the NLST and NELSON. As a positive screening resulted in a diagnostic work-up, the utilization of the diagnostic work-up could also be derived from these results. The unit cost of a LDCT in the NLST is obtained from the 'Physician Fee Application'.[41] The unit cost of a diagnostic-work up in the NLST is based on a study by Roth et al.[42] Roth used asymptomatic participants at high risk as the target population. Costs are based on the utilization of diagnostics. Tariffs are based on reimbursement rates by Medicare. The unit cost of a LDCT in the NELSON is an expert-based assumption. The unit cost for a diagnostic work-up in NELSON is a weighted average of all diagnostics used after a positive screen. The utilization of diagnostics in a diagnostic work-up in NELSON are derived from a study by Brinkhof et al.[43] The study collected data in real-world practice on the utilization of diagnostics in a population of people with clinical suspicion of lung cancer. The tariffs used in the NELSON are acquired from the 'DBC tariff application' hosted by the Dutch Healthcare Authority (NZa).[44] Utilization rates of some diagnostics are slightly adjusted, Because of a different target (Appendix - Table 7). Tariffs used both trials are from 2015. The tariffs in the NELSON are converted at an exchange rate of 1.12 euro to dollar.[45]

Model and assumptions

A decision tree is built in Microsoft Excel.[46] To represent the different modules of the screening strategies, as described before, the tree was also subdivided as such.

The screening algorithm is the structure of the decision tree. The screening algorithm consists of; different types of nodule measurement methods, nodule categories, number of screening rounds, and the duration of screening rounds.[17] Probabilities of screening results and diagnostic work-ups are

derived from study protocols (Appendix - Table 6).[2,3,17,47] Treatment of lung cancer is not included in the decision tree.

The parameters of the model are: the size of the population, the probability of developing lung cancer, the unit costs of the LDCT and diagnostic work-up, and survival rates. The probability of developing lung cancer is used to reflect the population characteristics. This probability is calculated by adding the probability of a false negative scan to the probability a true positive scan.

Analytical methods

To compare the cost-effectiveness of the screening approaches as evaluated in the trials, an incremental cost effectiveness ratio (ICER) is calculated by dividing the incremental costs by the incremental QALYs. Bootstrapping is used to simulate the uncertainty of the base case parameters. A total of 10,000 replicas using a uniform distribution are generated. An uniform distribution is used because of an unknown spread in all parameters. A willingness-to-pay (WTP) threshold of $100,000 per gained QALY is used.[48]

An one-way sensitivity analysis is used to determine the influence of the parameters in the decision tree. The parameters examined are: the probability of developing lung cancer, costs of the LDCT, costs of the diagnostic work-up, survival rates and discounting of utilities and costs.

A scenario analysis is used to determine the effect of substituting the algorithms of both trials. The substitution included the probabilities of screening and diagnostic results, and the probabilities of tumor stages. This scenario analysis could answer the question of what would happen if the trials would have used other measurement methods, other thresholds and an additional screening result category.

Information from the Fred Hutchinson lung cancer data registry will be analyzed to gather information on the generalizability of the NLST outcomes. Data from the registry is compared to the real-world participant characteristics with the trial participant characteristics. Collected data on age, gender, race, smoking history, smoking cessation, and time since quitting of the participants will be compared to characteristics of the NLST participants.

Results

Base case results

Model parameters can be found in Table 1. Life-years are calculated by the duration of the screening rounds, the probability of developing lung cancer, the weighted survival rates of the present tumors and the normal chance of dying of an average trial participant. Normal probability of dying is higher in the NLST (Appendix - Table 5).[17,35] Life-years collected in the NLST are the same in every year, because of the same duration of the screening rounds.

The utilities are assessed per tumor stage (Appendix - Table 4). Differences between gender are only detected in the ‘no cancer’ category. The utilities gained per screening result are calculated by using the gender distribution per study and the tumor stage distribution per gender (Appendix - Table 1 and

Appendix - Table 4). The tumor stage distribution per gender was published only by NELSON, the same distribution is used for the NLST.

		NLST	NELSON
Participants		26722	7915
Probability of developing lung cancer		2.56	2.73
Number of tumors	Stage I-II	481 (70%)	167 (77%)
	Stage III	116 (17%)	38 (18%)
	Stage IV	87 (13%)	10 (5%)
Life-years			
Screening result	Screening round		
Negative	1	0.995	0.998
	2	0.995	1.996
	3	0.995	2.494
Positive	1	0.904	0.904
	2	0.904	1.692
	3	0.904	2.058
Utilities			
Screening result			
Negative		0.7517	0.7565
Positive		0.6803	0.6676
LDCT costs ($)		186.89	224.00
Diagnostic work-up ($)		1270.00	2075.58

Table 1: Decision tree input

The outcomes of the base case analysis can be found in Table 2. The probability of developing lung cancer is higher in the NELSON. The percentage of tumors found in stage I/II is higher in the NELSON.

Variables	NLST (26,722)	NELSON (7,915)
Amount of screens	3	3.33
Amount of diagnostic work-ups	0.72	0.07
Life-Years	2.98	5.47
QALYs	2.14	3.92
Costs ($)	1,383.94	1,110.32

Table 2: Base case analysis per participant

ICER simulation

The parameters used can be found in Table 3 and Appendix - Table 8. The ranges in the probability of developing lung cancer are based on estimation of all trial information published. The range of the LDCT costs is based on unit costs in the 'Physician Fee Application'.[41] The range in diagnostic work-up costs in the NLST is adopted from the budget impact analysis of Roth et al.[42] The range in diagnostic work-up costs in NELSON is defined by adjusting the decrease of the CT and PET/CT to from 0-40% (Appendix -

Table 7). The five-year survival is a weighted average calculated by the distribution and survival rates of tumors in both trails (Appendix - Table 4).

	NLST			NELSON		
Parameters	Base case	Low	High	Base case	Low	High
Probability of developing lung cancer (%)	2.56	2.0	3.50	2.73	2.00	3.50
Cost of LDCT ($)	186.89	136.92	236.89	224.00	200.00	300.00
Cost of diagnostic work-up ($)	1,270.00	1,016.00	1,524.00	2,075.58	1,679.32	2,424.23
Weighted average of five-year survival	0.34		0.42	0.49	0.39	

Table 3: Parameter ranges ICER simulation

The original costs and QALYs of both trials in the base case are plotted in Figure 2. The NELSON protocol incurs less costs compared to NLST, and NELSON yet collected more QALYs.

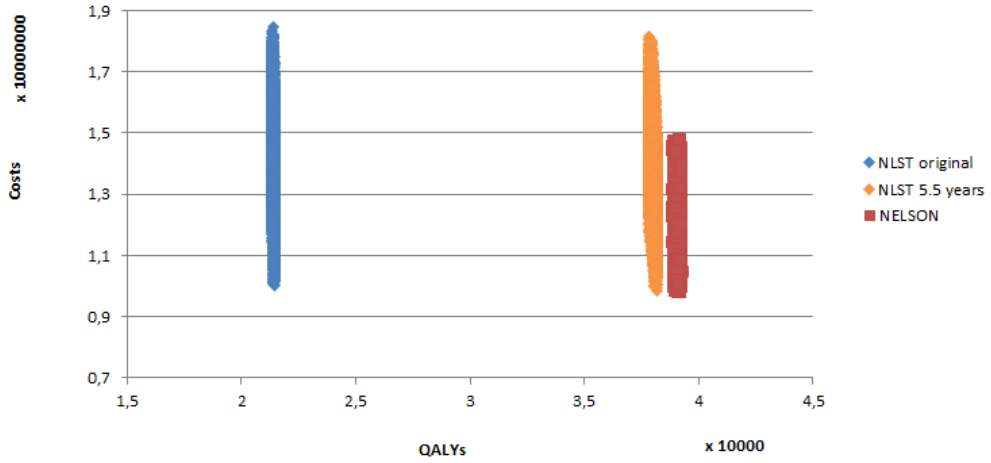


Figure 2: Costs and QALYs per 10,000 participants

The difference is QALYs is also because a different time span of the trials. To adapt for this difference, the third round of the NLST is stretched to 3.5 years. In this simulation, no extra screening or work-ups are performed. The total costs in the NLST decreased very slightly, because of a longer discounting period. The QALYs differ due to extra life-years collected in the stretched third round (Figure 2).

After stretching the third round of the NLST to 3.5 years, the NELSON population is still collecting more QALYs, and trial costs of NELSON are lower.

To include uncertainty, ICERs are calculated using bootstrapping (Figure 3). The x-axis contains the incremental QALYs collected by NELSON versus the NLST. The y-axis contains the incremental costs of the NELSON versus the NLST. In 78% of the simulated ICERs, NELSON gained less costs and more QALYs than the NLST (Appendix - Table 9). In the residual 22%, the NELSON has gained more QALYs than the NLST, but the costs are higher. Still, the cost-effectiveness is way below the threshold of $100,000 per QALY.

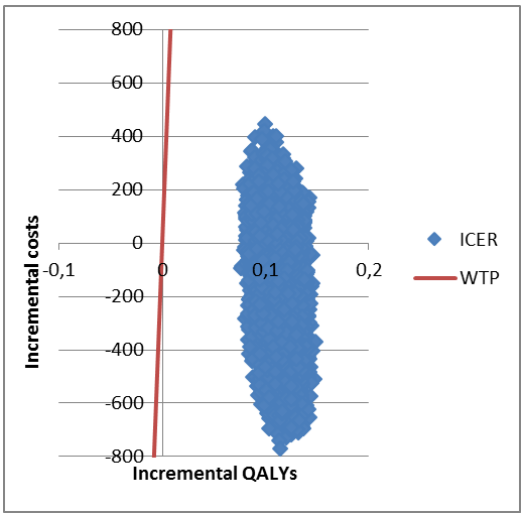


Figure 3: ICER scatterplot

One-way sensitivity analysis

Tornado diagrams are created to show the influence of the parameters on the cost-effectiveness. Ranges used in these diagrams can be found in Table 3 and Appendix - Table 8. Effects are determined in costs, QALYs and cost per QALY (Appendix - Table 10).

The unit cost of a LDCT scan, unit cost of diagnostic work-up, and probability of developing lung cancer are drivers of costs in the NLST (Appendix - Figure 1). None of the factors are a driving factor in terms of QALYs (Appendix - Figure 2). The costs per QALY are affected by both the cost parameters, and the probability of developing lung cancer (Figure 4).

Only the unit cost of a LDCT scan is a driver of the NELSON costs (Appendix - Figure 3). Changing utility discounting influences the total amount of QALYs collected (Appendix - Figure 4). The costs per QALY are mostly affected by the cost parameters, and the probability of developing lung cancer (Figure 4).

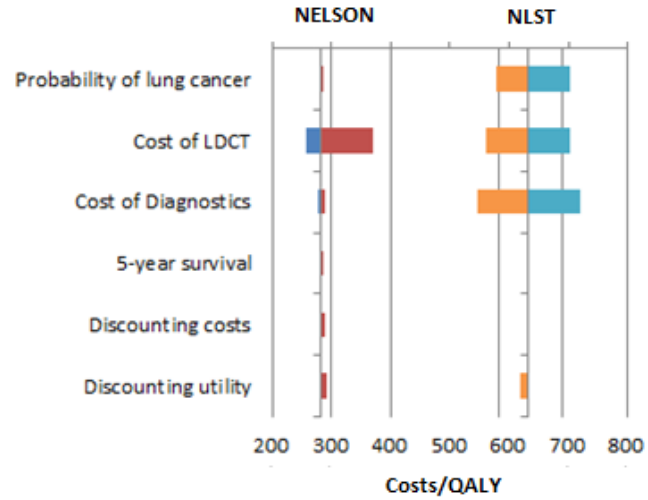


Figure 4: Parameter influence on costs per QALY

Scenario analysis

Scenario analysis is used to determine the effect of switching algorithms between the trials (Table 4). The difference due to switching algorithms can be seen between the brackets. The difference in tumors found is only due to the different algorithms.

The NLST algorithm causes less screens, and more life-years. The NELSON algorithm causes a significant lower amount of diagnostic work-ups, more QALYs and fewer costs.

Variables	NLST population NLST algorithm	NLST population NELSON algorithm	NELSON population NELSON algorithm	NELSON population NLST algorithm
Population size	26,722	26,722	7,915	7,915
Probability of lung cancer	2.56%	2.56%	2.73%	2.73%
Amount of lung cancers	684	684	216	216
Stage I-II	481	530 (10.19%)	167	152 (-8.98%)
Stage III	116	121 (4.31%)	38	37 (-2.63%)
Stage IV	87	33 (-62.07%)	10	28 (180.00%)
Number of screens	80,166	88,877 (10.87%)	26,325	23745 (-9.80%)
Number of diagnostic work-ups	19,260	1,757 (-90.88%)	533	5,888 (1004.69%)
Life-Years	144,728	145,708 (0.68%)	43,284	42,952 (-0.77%)
QALYs	101,456	101,060 (-0.39%)	31,019	31,147 (0.41%)
Costs ($)	36,428,838	23,997,432 (-34.13%)	8,788,208	16,204,572 (84.39%)

Table 4: Trial characteristics in per screening algorithm

Data registry analysis

The comparison of participant characteristics of the NLST and the lung cancer registry can be found in Table 5. The distribution of gender and race are similar. It is possible to join the real-world screening program if people are younger than 55 or never-smokers. As a result only 55% of the participants would be eligible for the actual NLST (Appendix - Figure 5). A big distribution of participants older than 65 compensates for the possibility to enter at younger age. The possibility to enter as a non-smoker can cause a healthier population in actual practice. This effect can also be caused by the bigger part of people who stopped smoking over 10 years ago. Altogether, the real-world screening population seems healthier than the NLST participants. If the real-world participants are healthier than the trial participants, the question arises if there has been a health volunteer bias. The better health of the real-world participants causes decrease in cost-effectiveness of the screening strategy in a real-world setting.

Participant characteristic	Category	Registry	NLST
Age	<55	15%	0%
	55-60	21%	43%
	60-65	21%	31%
	65-70	28%	18%
	70-75	12%	9%
	>75	3%	0%
Gender	Male	66%	59%
Race	Caucasian	92%	90%
	Non-hispanic black	4%	4%
	Hispanic	0%	2%
	Asian	0%	2%
	Other	3%	2%
Smoking history	Current	44%	48%
	Former	46%	52%
	Never	10%	-
Time since quitting	<5	26%	29%
	5-10	32%	33%
	>10	42%	38%

Table 5: Participant characteristics registry and NLST

Discussion & recommendations

The main objective of the study is to determine the influence of the population, algorithm and diagnostic work-up on the cost-effectiveness of the screening strategy, and finding the drivers of cost-effectiveness in the NLST and NELSON. The secondary objective is to determine the generalizability of the NLST's screening strategy by using real-world practice data.

Scenario analysis shows that the algorithm seems by far the most important cause of difference in cost-effectiveness between the trials. When the NLST algorithm was used in the NELSON setting, the number of screens would decrease with 9.80%, however the number of diagnostic work-up would increase enormously with more than a ten-fold. Using the NELSON algorithm in the NLST setting would increase the number of screens with 10.87%, but decrease the amount of diagnostic work-ups with 90.88%. Switching algorithms had almost no effect on QALYs, but the NELSON algorithm in the NLST setting would decrease costs with 34.13%, and the NLST algorithm in the NELSON setting would increase the cost by 84.39%.

The simulation of the ICERs shows that the NELSON screening strategy is cost-effective compared to the NLST screening strategy. The NELSON strategy was dominant over the NLST in the majority (78.7%) of the simulation, and cost-effective in the remaining part (21.3%) using a $100,000 WTP threshold. The one-way sensitivity analysis shows that the NLST strategy is driven by the unit cost of the LDCT, unit cost of the diagnostic-work up, and the probability of developing lung cancer. The only driver in the NELSON trial is the unit cost of the LDCT.

The data registry analysis shows that participants are healthier in real-world practice than in the NLST. More screenings are needed to detect a tumor in a healthier population. The cost-effectiveness study by Black showed that the cost-effectiveness decreased when population is healthier.[23] Besides, the analysis of the data in the registry showed a large amount (82%) of participants with incidental findings (Appendix - Figure 7).[30] Published data by the NELSON and NLST showed 8-10% of incidental findings in the trials.[47,49] Black suggests that cost-effectiveness will decrease if there is an increase in incidental findings.[23] The exact effect of incidental findings on the cost-effectiveness is unknown.

The use of trial and real-world data is one of the strengths of this study. Other strengths are the different types of analysis cause a multi-perspective view on the cost-effectiveness between the trials and the drivers within each trial.

In order to compare the NLST and NELSON, it was necessary to use the same time span in both trials. The NELSON used different durations per screening round, to do research on the influence of the time span. The set-up of the NELSON is used like it would be implemented in real-world practice in that format. The choice of expanding the third round of the NLST to 3.5 years is of great impact on the outcome of the study. Another choice would be to stop NELSON after two screening rounds, at which NELSON would have also had a time span of 3 years. Stopping after two rounds would favor the NLST screening strategy (Appendix - Figure 6). Instead of stretching the third round to 3.5 years, another option would be an addition of a fourth and fifth screening round, both with a one-year span. The addition would probably cause a higher amount of lung cancers (so an unhealthier population in the model) and would cause a big increase in costs due to all additional screenings and diagnostic work-ups. An unhealthier population is already simulated, and the extra costs that would be made in the two additional rounds would only be in favor of the NELSON. So modeling a fourth and fifth round would not be in favor of the NLST. Therefore, the choice to expand the third round of the NLST is made in the belief that, for the model used in this study, it would simulate actual practice the most. The stretch of the third NLST round to 3.5 years simulated no use of screenings or diagnostics. This caused no increase in costs.

The choice to use different discount rates for the trials is based on the background of the study. The study is not based on determining which screening strategy is better, but to assess which elements are of big influence on the cost-effectiveness, so that screening strategies can improve with the focus on elements that matter the most.

The characteristics of the population are only simulated by the probability of developing lung cancer. Characteristics that can give more information about the population are the minimum and maximum age, smoking cessation and pack-years. To add these characteristics in the model, patient-level data needs to be available, or research on relative risks on all of these characteristics needs to be carried out. Still, even if this information is known, implementing the characteristics in the model can be tough, because other relevant factors like health insurance and comorbidity of patients can be of influence and are harder to concretize.

The unit costs for a diagnostic work-up are derived from different populations, not only in a geographic sense. The unit cost in NELSON is based on a population in clinical suspicion of lung cancer. The unit cost

in the NLST is based on an asymptomatic population at high risk of developing lung cancer. The difference in unit costs in the NLST and NELSON can be an underestimation as well as an overestimation of the actual difference. The simulation used increased and decreased unit costs for both trials.

The probability of developing lung cancer decreases if the population gets healthier. More QALYs will be gained by the population. The ICER can possibly stay intact if both trials experienced a healthier population. However, the ICER would change in comparison to a 'no-screening' setting. This could possibly be in favor of the 'no-screening' setting. This information is not processed in the decision tree, due to missing information of control groups in both studies.

This is the first study to compare the cost-effectiveness of the NLST and NELSON. The cost-effectiveness of the NLST screening program in comparison to no screening is already determined in a study by Black et al.[50] The conclusion from that study was that the NLST is cost-effective compared to no screening, but that the verdict heavily depends on assumptions made in the study. If the NLST would be cost-effective compared to no-screening, this doesn't implicate that the NELSON would be cost-effective compared to no-screening. More information on the NELSON control-arm is needed to conclude if the NELSON trial is cost-effectiveness in comparison to no-screening. If the NELSON control-group relates the same to the trial participants as the NLST control and trial do, NELSON would be cost-effective to no-screening.

This study showed that the NELSON is cost-effective compared to the NLST. Knowing the cost-effectiveness of trials is useful; however, translating this to cost-effectiveness in actual practice is of more value. A study by NELSON proved that 16.6% of the trial participants quit smoking, which is much higher than the 3-7% in general population.[51] The cessation of smoking is still the best way to prevent cancer, but this effect causes a healthier screening population than expected, which causes a lower cost-effectiveness. Goulart states that the cost-effectiveness of the screening program in actual practice will depend largely on the implementation of the program.[52] He points to four main factors that are of big influence: (1) generalizability of the population, (2) the effect of false-positive screening results, (3) the expertise of screening centers, and (4) the infrastructure of screening programs.

This study showed that the population of the NLST is not generalizable to a real-world US screening population, due to differences in age, smoking history and incidental findings. To made a conclusion about the cost-effectiveness in an actual setting, more research needs to be carried out on the other three factors pointed out by Goulart. The study showed that the algorithm of the NELSON seems more cost-effective. The biggest difference in algorithms is the addition of the 'indeterminate' category. The study also showed that the probability of lung cancer is a driver in the NLST, but not in the NELSON. Accepting healthier people in the NELSON would not result in a big difference in cost-effectiveness, meanwhile a different population in the NSLT would influence the cost-effectiveness massively. Recommendations to improve the cost-effectiveness of the NLST would be to focus on the eligibility of the population and adjusting the screening algorithm.

References

1. Humphrey, L. L. *et al.* Screening for Lung Cancer With Low-Dose Computed Tomography: A Systematic Review to Update the U.S. Preventive Services Task Force Recommendation. *Ann Intern Med* **159,** 411–420 (2013).

2. Aberle, D. R. *et al.* Results of the two incidence screenings in the National Lung Screening Trial. *N. Engl. J. Med.* **369,** 920–931 (2013).

3. The National Lung Screening Trial Research Team. Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. *BMJ Br. Med. J.* **365,** (2011).

4. Medicare. Your Medicare Coverage: Lung cancer screening. (2015). at <http://www.medicare.gov/coverage/lung-cancer-screening.html>

5. NCCN. Lung Cancer Screening. *Version I.2012* (2012). at <nccn.org>

6. Andermann, A., Blancquaert, I., Beauchamp, S. & Déry, V. Revisiting Wilson and Jungner in the genomic age: A review of screening criteria over the past 40 years. *Bull. World Health Organ.* **86,** 317–319 (2008).

7. Cipriano, L., Romanus, D. & Earle, C. Lung cancer treatment costs, including patient responsibility, by stage of disease and treatment modality, 1992–2003. *Value Heal.* **14,** 41–52 (2011).

8. Howlader, N. *et al.* Cancer statistics review 1975-2010. *J. Natl. Cancer Inst.* (2011).

9. IKNL. Overleving Longkanker. (2015). at <http://www.cijfersoverkanker.nl/selecties/dataset_1/img54d89a79332ab>

10. The American Cancer Society. Non-small cell lung cancer survival rates by stage. (2007). at <http://www.cancer.org/cancer/lungcancer-non-smallcell/detailedguide/non-small-cell-lung-cancer-survival-rates>

11. Hackshaw, A. K., Law, M. R. & Wald, N. J. The accumulated evidence on lung cancer and environmental tobacco smoke. *BMJ* **315,** 980–988 (1997).

12. U.S. Department of Health and Human Services. *The Health Consequences of Involuntary Exposure to Tobacco Smoke A Report of the Surgeon General*. *Health San Francisco* 727 (2006). at <http://www.surgeongeneral.gov/library/reports/secondhandsmoke/index.html>

13. Anthonisen, N. R. *et al.* The effects of a smoking cessation intervention on 14.5-year mortality. *Ann. Intern. Med.* **142,** 233–239 (2005).

14. The American Cancer Society. Cancer Facts & Figures: 2014. *Cancer Facts Fig.* (2014). at <http://www.cancer.org/acs/groups/content/@research/documents/webcontent/acspc-042151.pdf>

15. Mariotto, A. B., Robin Yabroff, K., Shao, Y., Feuer, E. J. & Brown, M. L. Projections of the cost of cancer care in the United States: 2010-2020. *J. Natl. Cancer Inst.* **103,** 117–128 (2011).

16. RIVM. Costs of lung cancer. (2011). at <http://www.kostenvanziekten.nl/systeem/kosten-van-ziekten-tool/Default.aspx?ref=kvz_v2l1b1p4r3c7i0t1j0o3y6a-1g0d202s54z0f0w2>

17. Horeweg, N. Lung cancer screening in the NELSON trial: balancing harms and benefits. (Erasmus University Rotterdam, 2014).

18. Pedersen, J. H. *et al.* The Danish randomized lung cancer CT screening trial--overall design and results of the prevalence round. *J. Thorac. Oncol.* **4,** 608–614 (2009).

19. Becker, N. *et al.* Randomized study on early detection of lung cancer with MSCT in Germany: Study design and results of the first screening round. *J. Cancer Res. Clin. Oncol.* **138,** 1475–1486 (2012).

20. Baldwin, D. R. *et al.* UK Lung Screen (UKLS) nodule management protocol: modelling of a single screen randomised controlled trial of low-dose CT screening for lung cancer. *Thorax* **66,** 308–313 (2011).

21. Pastorino, U. *et al.* Annual or biennial CT screening versus observation in heavy smokers. *Eur. J. Cancer Prev.* **21,** 308–315 (2012).

22. Infante, M. *et al.* Lung cancer screening with spiral CT. Baseline results of the randomized DANTE trial. *Lung Cancer* **59,** 355–363 (2008).

23. Black, W. C. *et al.* Cost-effectiveness of CT screening in the National Lung Screening Trial. *N. Engl. J. Med.* **371,** 1793–1802 (2014).

24. National Lung Screening Trial Research Team. The National Lung Screening Trial: Overview and Study Design. *Radiology* **258,** (2010).

25. Winkleby, M. A., Jatulis, D. E., Frank, E. & Fortmann, S. P. Socioeconomic status and health: How education, income, and occupation contribute to risk factors for cardiovascular disease. *Am. J. Public Health* **82,** 816–820 (1992).

26. Mackenbach, J. P. *et al.* Socioeconomic inequalities in health in 22 European countries. *N. Engl. J. Med.* **358,** 2468–2481 (2008).

27. American Psychological Association. Education and Socioeconomic Status. (2015). at <http://www.apa.org/pi/ses/resources/publications/factsheet-education.aspx>

28. CDC. Health Disparities and Inequalities Report - United States, 2011. *Morb. Mortal. Wkly. Rep.* **60,** (2011).

29. U.S. Department of Health and Human Services. Health, United States, 2011. *DHHS* (2012). doi:10.1080/01621459.1987.10478476

30. Fred Hutchinson Cancer Research Center. Lung Cancer Screening Data Registry. (2015).

31. Goulart, B. H. L. Consent to take part in a research study: Performance of Low-Dose Computed Tomography for Early Detection of Lung Cancer : a Pilot Registry Study. **Version 3.,** (2014).

32. Severens, J. L. & Milne, R. J. Discounting Health Outcomes in Economic Evaluation: The Ongoing Debate. *Value Heal.* **7,** 397 (2004).

33. Van Bortel, L., Vanmolkot, F. & Van Lancker, G. *Handboek farmaceutische geneeskunde*. 257 (2014).

34. IKNL. Overleving Long en Luchtpijpkanker. (2015).

35. Aberle, D. R. *et al.* Baseline characteristics of participants in the randomized national lung screening trial. *J. Natl. Cancer Inst.* **102,** 1771–1779 (2010).

36. CBS. Overlijdensratio 55-60 Nederland. (2014).

37. CDC. National Vital Statistics Report: Deaths: Final Data for 2013. **64,** (2014).

38. Black, W. C. *et al.* Cost-effectiveness of CT screening in the National Lung Screening Trial: Supplementary Appenxi. *N. Engl. J. Med.* **371,** 1793–1802 (2014).

39. Brazier, J., Roberts, J. & Deverill, M. The estimation of a preference-based measure of health from the SF-36. *J. Health Econ.* **21,** 271–292 (2002).

40. Gareen, I. F. *et al.* Impact of lung cancer screening results on participant health-related quality of life and state anxiety in the National Lung Screening Trial. *Cancer* 3401–3409 (2014). doi:10.1002/cncr.28833

41. Centers for Medicare and Medicaid Services. Physician Fee Application. (2015). at <http://www.cms.gov/apps/physician-fee-schedule/search/search-results.aspx?Y=13&T=0&HT=0&CT=3&H1=31628&M=5>

42. Roth, J. A. *et al.* Projected Clinical, Resource Use, and Fiscal Impacts of Implementing Low-Dose Computed Tomography Lung Cancer Screening in Medicare. (2015).

43. Brinkhof, S., Groen, H. J. M., Siesling, S. & Ijzerman, M. J. *Health Economic and Budget Impact of Lung Cancer Diagnostic Procedures*. (2014).

44. Dutch Healthcare Authority. DBC Tariff Application. (2015). at <http://dbc-zorgproducten-tarieven.nza.nl/nzaZpTarief/ZoekfunctieDot.aspx>

45. Bloomberg. EURUSD Spot Exchange Rate. (2015). at <http://www.bloomberg.com/quote/EURUSD:CUR>

46. Microsoft. Excel. (2010).

47. The National Lung Screening Trial Research Team. Results of Initial Low-Dose Computed Tomographic Screening for Lung Cancer. *N Engl J Med* **368,** 1980–1991 (2013).

48. Neumann, P. J., Cohen, J. T. & Weinstein, M. C. Updating Cost-Effectiveness — The Curious Resilience of the $50,000-per-QALY Threshold. 2014–2015 (2014). doi:10.1056/NEJMp1405158

49. Van de Wiel, J. C. M. *et al.* Neglectable benefit of searching for incidental findings in the Dutch-Belgian lung cancer screening trial (NELSON) using low-dose multidetector CT. *Eur. Radiol.* **17,** 1474–1482 (2007).

50. Black, W. C. Computed tomography screening for lung cancer in the National Lung Screening Trial: A Cost-effectiveness Analysis. *J Thorac Imaging* **30,** 79–87 (2015).

51. Van der Aalst, C. M., van den Bergh, K. A. M., Willemsen, M. C., de Koning, H. J. & van Klaveren, R. J. Lung cancer screening and smoking abstinence: 2 year follow-up data from the Dutch-Belgian randomised controlled lung cancer screening trial. *Thorax* **65,** 600–605 (2010).

52. Goulart, B. H. L. The Value of Lung Cancer CT Screening : It Is All about Implementation. *Asco Educ. B.* (2015).

Appendix

Tables

Appendix Tables

	Category	NLST	NELSON
Participant characteristic			
Number of participants		26722	7915
Age	<55	0%	
	55-59	43%	
	60-64	31%	
	65-69	18%	
	70-74	9%	
	>74	0%	
	Median		58
Gender	Male	59%	83%
Smoking status	Current	48%	56%
	Former	52%	44%
Race or ethnic	White	90%	
	Black	4%	
	Hispanic or Latino	2%	
	Asian	2%	
	Other	2%	
Pack-years	Median	48.0	37.8
Inclusion criteria			
Age	Minimum	55	50
	Maximum	74	75
Smoking history	Pack-years	≥ 30	≥ 15/day for 25 years / ≥ 10/day for 30 years
Smoking cessation	Time since quitting	≤ 15 years	≤ 10 years
Exclusion criteria			
Measurement of health	Self-reported	-	Moderate or bad
Activities	Ability to climb stairs	-	≤ 2 flights
Body weight		-	140 kg
	Unexplained body weight loss	≥ 6.8 kg	-
Other diseases	History of:	• Lung cancer • Removal of portion of the lung • Treatment of skin cancer ≤ 5 years prior to eligibility assessment • Hemoptysis • Pneumonia or acute respiratory	• Lung cancer under treatment or diagnosed ≤ 5 years prior to eligibility assessment • Renal cancer • Melanoma • Breast cancer

		infection treated with antibiotics ≤ 12 weeks to eligibility assessment	
Chest CT examination	Time since last examination	≤ 18 months	≤ 12 months
Implants or devices	Metallic implants or devices	Chest or back	-
Requirement for		Home oxygen supplementation	-
Participation in other studies		Cancer prevention study other than smoking cessation study	-

Appendix - Table 1: Participants characteristics NLST and NELSON

Participant characteristic	Category	
Age	<55	15%
	55-59	21%
	60-64	21%
	65-69	28%
	70-74	12%
	>74	3%
Gender	Male	66%
Smoking status	Current	44%
	Former	46%
	Never	10%
Race or ethnic	White	92%
	Black	4%
	Hispanic or Latino	0%
	Asian	0%
	Other	3%
Pack-years	Median	35.0

Appendix - Table 2: Participant characteristics data registry

Study	Measurement method	Thresholds	Result
NLST	Diameter	<4 mm	Negative
		>4 mm	Positive
NELSON	Diameter	<5 mm	Negative
		>5 mm to <10 mm	Indeterminate
		>10 mm	Positive
	Volume	<100 mm3	Negative
		>100 mm3 to <300 mm3	Indeterminate
		>300 mm3	Positive
	Volume Doubling Time	>600 days	Negative
		<600 days to >400 days	Indeterminate
		<400 days	Positive

Appendix - Table 3: Categorization thresholds

Stage	% Men NELSON	% Women NELSON	Survival rate US	Survival rate the Netherlands	Utilities
No cancer					0.76 (men) 0.74 (women)
IA	62	85	0.49	0.66	0.72
IB	6	3	0.45	0.42	0.695
IIA	8	-	0.30	0.33	0.67
IIB	-	-	0.31	0.15	0.67
IIIA	15	12	0.14	0.08	0.645
IIIB	4	-	0.05	0.03	0.645
IV	6	-	0.01	0.00	0.62

Appendix - Table 4: Population characteristics per stage

Age-group	United States	The Netherlands
55-59	0.007	0.005
60-64	0.010	0.008
65-69	0.015	0.012

Appendix - Table 5: Probability on dying per age group

Screening round	Type	Result	NLST	NELSON
1	Screening	Positive	24.2	1.6
		Negative	75.8	79.2
		Indeterminate	-	19.2 (94.7% neg)
	Diagnostic work-up	False positive	96.2	66.3
		False negative	0.1	0.1
2	Screening	Positive	27.3	1.2
		Negative	72.7	92.2
		Indeterminate	-	6.6 (91.5% neg)
	Diagnostic work-up	False positive	97.6	56.2
		False negative	0.1	0.3
3	Screening	Positive	27.9	1.3
		Negative	72.1	91.9
		Indeterminate	-	6.8 (83.9% neg)
	Diagnostic work-up	False positive	94.8	54.4
		False negative	0.1	0.2

Appendix - Table 6: Screening results per round

Modality	Probability (%)	NZa Tariff code	Tariff (€)	Decrease due to population (%)	Weighted cost (€)
X-thorax	87.1	85002	55.1		48.34
X-thorax (2nd)	12.9				7.17
PET/CT	79.1	120501	1176.53	10.00	837.79
PET/CT (2nd)	2.2			10.00	23.27
CT	97.8	86042	215.52	10.00	189.70
CT (2nd)	5.5%			10.00	10.66
Cervical mediastinoscopy	5.8	83190	246.06		14.20
EUS/EBUS	26.9	34386	765.76		206.16
EUS/EBUS (2nd)	1.1				8.41
MR Brain	23.1	81092	253.89		58.59
Bone scintigraphy	3.0	80080	109.09		3.30
US guided pleura punction	13.5	32684	51.08		6.88
Bronchoscopy	66.8	32480	374.97		250.32
Bronchoscopy (2nd)	8.8				32.96
X-ray guidance bronchoscopy	15.7	85000	52.79		8.27
CT guided biopsy	15.4	80047	200.11		30.79
X-thorax	9.9	85002	55.51		5.49
VO2 max	11.0	39844	148.01		16.26
Broad lung function	56.3	39932	78.23		44.06
Broad lung function (2nd)	0.5				0.43
Flow volume	9.6	39839	42.91		4.13
Flow volume (2nd)	0.5				0.24
Ventilation/perfusion scan	8.2	120060	228.44		18.83
ECG	61.3	39757	43.99		26.95
KRAS/EGFR	28.8	50512	454.53	100.00	-
ALK	6.3	50514	985.07	100.00	-

Appendix - Table 7: Diagnostic resource utilization NELSON

Number of tumors						
	NLST			NELSON		
Stage	Base case	Low	High	Base case	Low	High
IA	329	299	359	137	122	152
IB	71	51	91	11	5	17
IIA	26	18	34	14	8	20
IIB	20	14	26	0	0	5
IIIA	59	44	74	30	22	38
IIIB	49	38	60	7	2	12
IV	81	60	102	10	4	16
Utilities						
Stage	Base case	Low	High			
IA	0.72	0.70	0.74			
IB	0.695	0.68	0.71			
II	0.67	0.65	0.69			
III	0.645	0.63	0.66			
IV	0.62	0.60	0.64			
Discounting						
Variable	NLST			NELSON		
Discounting costs	3	-	4	4	3	-
Discounting utilities	3	1.5	-	1.5	-	3

Appendix - Table 8: Ranges in simulation

	NLST (%)	NELSON (%)
Dominance	-	78.04
Cost-effective but no dominance	-	21.96

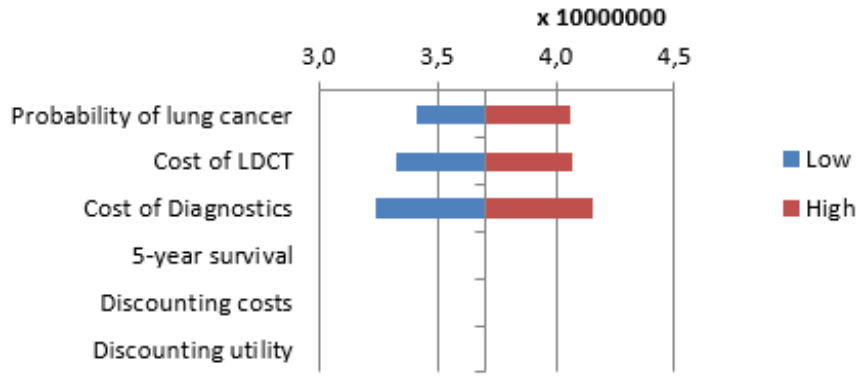
Appendix - Table 9: ICER bootstrapping result

Parameter	Effects on	NLST		NELSON	
Ranges		Low	High	Low	High
Probability lung cancer	Costs	7.69	9.78	1.15	1.03
	QALYs	0.15	0.10	0.13	0.12
	Cost/QALY	7.78	9.94	1.27	1.16
Cost of LDCT	Costs	10.12	10.12	8.55	30.25
	QALYs	-	-	-	-
	Cost/QALY	10.12	10.12	9.55	30.25
Cost of diagnostic work-up	Costs	12.43	12.43	2.07	1.82
	QALYs	-	-	-	-
	Cost/QALY	12.43	12.43	2.07	1.82
Discounting costs	Costs	1.80	-	-	2.23
	QALYs	-	-	-	-
	Cost/QALY	1.80	-	-	2.23
Five-year survival	Costs				
	QALYs	-	0.26	0.31	-
	Cost/QALY	0.26	-	-	0.31
Discounting utility	Costs				
	QALYs	-	2.21	3.23	-
	Cost/QALY	2.16	-	-	3.34

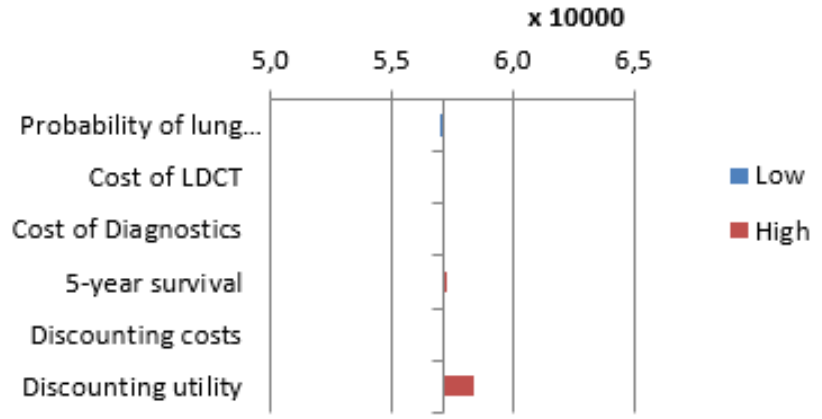
Appendix - Table 10: Effects on costs, QALYs and cost per QALY

Figures

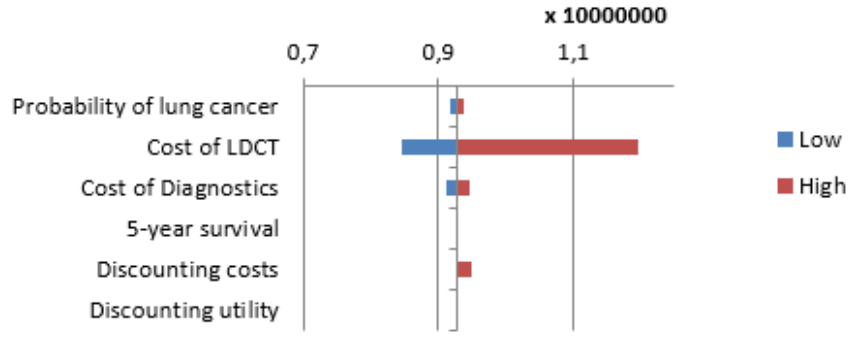
Figures in appendix



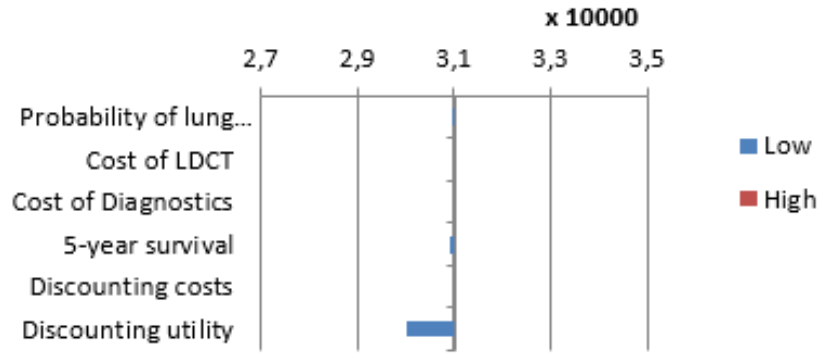
Appendix - Figure 1: Parameter influence on costs NLST



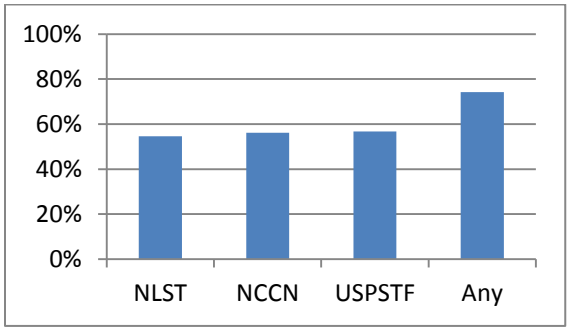
Appendix - Figure 2: Parameter influence on QALYs NLST



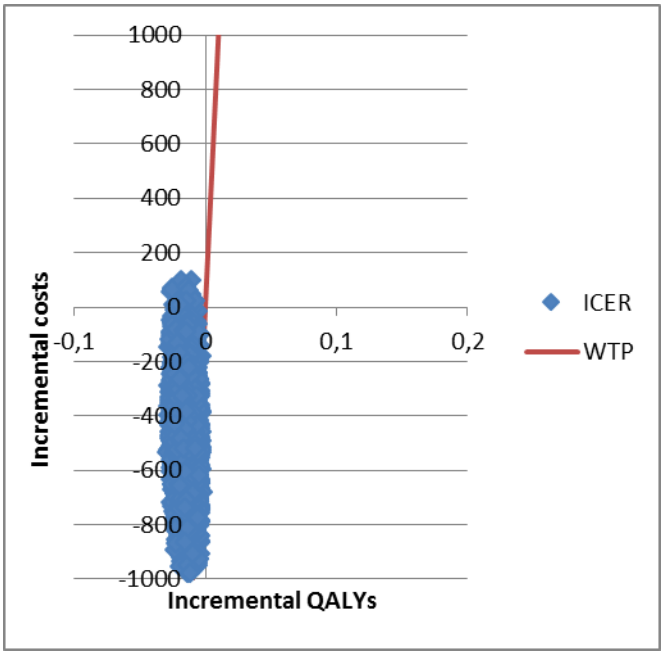
Appendix - Figure 3: Parameter influence on costs NELSON



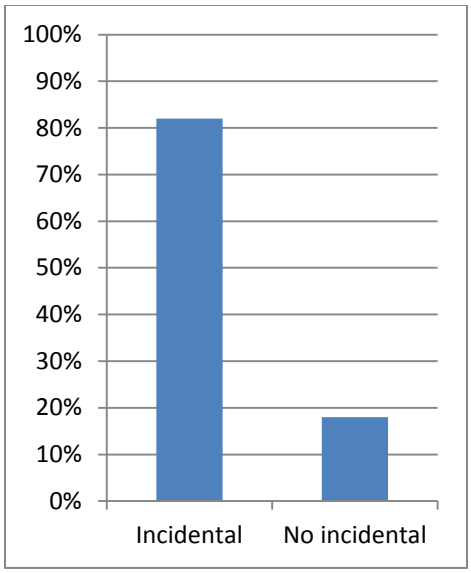
Appendix - Figure 4: Parameter influence on QALYs NELSON



Appendix - Figure 5: Participant eligibility in registry



Appendix - Figure 6: ICER scatterplot if NELSON stopped after two screening rounds



Appendix - Figure 7: Incidental findings in data registry