VALUE IN HEALTH 21 (2018) 1269–1277

Available online at www.sciencedirect.com

ScienceDirect

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

Comparative-Effectiveness Research/HTA

Assessing Lung Cancer Screening Programs under Uncertainty in a Heterogeneous Population

Henk Broekhuizen, PhD[1,*], *Catharina G.M. Groothuis-Oudshoorn, PhD*[2], *Rozemarijn Vliegenthart, PhD*[3], *Harry J.M. Groen, PhD*[4], *Maarten J. IJzerman, PhD*[1,5]

[1]*Radboud University Medical Center, Department of Health Evidence, Nijmegen, The Netherlands;* [2]*University of Twente, Faculty of Behavioural Management and Social Sciences, Technical Medical Centre, Department of Health Technology and Services Research, Enschede, The Netherlands;* [3]*University of Groningen, University Medical Center Groningen, Department of Radiology, Groningen, The Netherlands;* [4]*University of Groningen, University Medical Center Groningen, Department of Pulmonary Diseases, Groningen, The Netherlands;* [5]*University of Melbourne, Faculty of Medicine, Dentistry and Health Sciences and Victorian Comprehensive Cancer Centre, Melbourne, Australia*

ABSTRACT

Background: Lung cancer screening can reduce cancer mortality. Most implementation studies focus only on low-dose computed tomography (LDCT) and clinical attributes of screening and do not include preferences of potential participants. In this study we evaluated the perceived value of screening programs based on LDCT, breath analysis (BA), or blood biomarkers (BB) according to the perspective of the target population. **Methods:** A multi-criteria decision analysis framework was adopted. The weights of seven attributes of screening (sensitivity, specificity, radiation burden, duration of screening process, waiting time until results are communicated, location of screening, and mode of screening) were obtained from an earlier study that included a broad sample from the Netherlands. Performance data for the screening modalities was obtained from clinical trials and expert opinion. Parameter uncertainty about clinical performances was incorporated probabilistically, while heterogeneity in preferences was analyzed through subgroup analyses. **Results:** The mean overall values were 0.58 (CI: 0.57 to 0.59), 0.57 (CI: 0.56 to 0.59), and 0.44 (CI: 0.43 to 0.45) for BB, BA, and LDCT, respectively. Seventy-seven per cent of respondents preferred BB or BA. For most subgroups, the overall values were similar to those of the entire sample. BA had the highest value for respondents who would have been eligible for earlier screening trials. **Discussion:** BB and BA seem valuable to participants because they can be applied in a primary care setting. Although LDCT still seems preferable given its strong and positive evidence base, it is important to take non-clinical attributes into account to maximize attendance.

Keywords: lung cancer screening, multicriteria decision analysis, public preferences, subgroup analysis, uncertainty.

Introduction

Although the prognosis of cancer survival has greatly improved over the last few decades, lung cancer mortality has not been significantly reduced, with the 5-year survival being only 18% [1]. Because the primary reason for the high mortality is that the disease has already metastasized once detected, it is found that early detection can substantially improve prognosis [2]. In the National Lung Screening Trial (NLST) it was shown that annual screening with low-dose computed tomography (LDCT) can reduce cancer-related mortality by 20% compared with annual screening with chest x-ray [3]. Given these promising results, policymakers across the world are considering whether lung cancer screening should be adopted as part of national screening programs [4]. Nevertheless, there is still no consensus about actual screening guidelines or how these might be implemented [5]. A few studies have been performed investigating how lung cancer screening using LDCT should be implemented, yet these focus only on optimizing clinical parameters such as screening frequency or eligibility criteria [6,7]. Furthermore, although there is considerable experience with CT-based screening, recent advances in the fields of breath analysis (BA) and blood biomarker (BB) screening modalities should be considered as well, because they might be preferred more by the general public because of convenience benefits [8–10]. Also, results from lung cancer screening trials suggest that there are some subgroups that are less likely to participate in screening after being invited compared with other subgroups [11,12]. This may be explained from previous research that shows that a decision to attend also depends on nonclinical attributes of screening such as comfort, duration, and location [13] or on person-specific factors such as self-perceived risk of lung

[*] *Address correspondence to:* Henk Broekhuizen, Radboud University Medical Center, Post 133, Department for Health Evidence, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands.

E-mail: henk.broekhuizenversteeg@radboudumc.nl.

1098-3015$36.00 – see front matter Copyright © 2018, ISPOR–The Professional Society for Health Economics and Outcomes Research. Published by Elsevier Inc.

https://doi.org/10.1016/j.jval.2018.01.021

cancer [14]. Because low screening attendance rates may lead to lower cancer-related mortality reductions than predicted in trials, a lung cancer screening guideline should be optimized for attendance and thus appreciate both clinical diagnostic performance and other factors related to the process and convenience of the modality [13].

Several survey methods such as multicriteria decision analysis (MCDA) or discrete choice experiments can be used to assess the relative value of hypothetical screening programs with different clinical and nonclinical attributes [15–17]. Studies using preference elicitation methods have been carried out in breast [18], colorectal [19–23], and prostate [24–26] cancers. The underlying assumption in these stated preference studies is that the screening program with the highest relative value is most preferred by the respondents, and therefore has the highest likelihood that the eligible population will attend. Although such preferences cannot directly predict screening attendance, they do provide insight into what attributes of screening programs influence the intention to attend. In a previous study, using a stated preference survey we identified seven attributes of screening programs for lung cancer that were most important for potential participants in the Netherlands [27]. The aim of the present study was to use MCDA to construct and evaluate the relative values for different screening programs that use LDCT, BA, or BB. A second aim was to investigate whether the value of screening programs differs between preference subgroups or risk profile subgroups.

Methods

MCDA is a methodology stemming from operations research that can support a wide range of decisions. In general, an MCDA distinguishes four consecutive steps: structuring, weighting, scoring, and aggregating [28,29]. During the structuring step, the decision problem needs to be defined (who is the decision maker and what are the decision options) and important attributes of the decision alternatives are selected. The relative importance or attribute weight of each included attribute is then assessed using elicitation methods such as swing weighting or MACBETH [28,30]. After this, the performance of each decision option on each attribute is measured and valued numerically in the scoring step. Then, the weights and scores are aggregated to obtain an overall value estimate for each decision option, and the decision option with the highest value is considered the preferred choice. In the final step of the MCDA process, the impact of various types of uncertainty in the scores, weights, and overall values should be investigated and the results of the MCDA should then be communicated to stakeholders.

Structuring the Decision

In this study, we investigated the decision whether to attend the first round of a screening program for lung cancer. We assumed the decision perspective of (potential) participants from a broadly determined eligible population in the Netherlands, which we assumed were all men and women aged between 40 and 80 years with no history of lung cancer. We assumed that someone is most likely to attend the screening program with the highest perceived value. Therefore, we aimed to assign a value to each alternative screening program using the respondents from a previously conducted stated preference study [27]. The program with the highest mean value across the respondent sample would be denoted as the preferred program. The screening programs in this study were based on three different modalities, namely, LDCT, BA, and BBs. In case of LDCT, a malignancy is suspected if the diameter or volume (growth) of a nodule on a low-dose thorax image exceeds a predefined threshold [5]. We focused on LDCT with a threshold based on volume-doubling time instead of a diameter-based malignancy threshold because a high false-positive rate was reported for the latter [3,31]. Regarding the second modality, BA, there are multiple technologies [32]. We chose to focus on the electronic nose ("e-Nose") technology [8]. With BA, one must exhale into a device similar to that used for lung function testing. The e-Nose detects (patterns of) volatile organic compounds in the exhaled air, which are then compared with patterns found in patients with lung cancer and healthy persons. If the chemical patterns in the collected breath sample match those of patients with lung cancer, the participant is referred for further testing. For the last modality, a multitude of BB-based screening technologies exist [9,33,34]. For the present study, we chose to focus on circulating tumor cells in blood [35]. With these so-called liquid biopsies, blood is taken from the participant and a laboratory procedure is used to find malignant cells that have shed into the vascular system from the primary tumor.

We compared the three screening modalities on seven attributes of screening that were identified in a previous study [27]. The attributes that were identified together with interviewees and panel members were sensitivity, specificity, radiation burden, duration of screening, waiting time until results are communicated, location of screening, and mode of screening. *Sensitivity* was defined as the probability of a positive (i.e., suspect) test result given that someone has lung cancer. *Specificity* was defined as the probability of a negative (i.e., nonsuspect) test result given that someone does not have lung cancer. *Radiation burden* was defined as the radiation that a participant receives during a single screening on top of the normal yearly background radiation in the Netherlands (in millisievert [mSv]) [36]. *Duration of screening* was defined as the time the participant would spend at the facility where screening takes place (in minutes). *Waiting time until results* was defined as the time until the participant receives the screening results (in days), either through a consultation or with a letter sent to his or her home address. *Mode of screening* was the modality of screening, with "going through a scanner," "exhaling into a device," and "giving a blood sample" as levels. The *location of screening* attribute had two levels: "the office of your general practitioner (GP)" and "the hospital nearest to you." For each attribute, a lower (worst) level and an upper (best) level were identified. The actual performance of the screening programs should fall in this range. Furthermore, it was assumed that respondents are indifferent to performance changes outside of this range; for example, performing worse than the lower level confers the same value as performance exactly at the lower level. For the sensitivity, specificity, and radiation burden attributes, the lower and upper levels were identified on the basis of literature and the clinical expertise of two of the authors.

In the preference survey, the percentages used for sensitivity and specificity were rounded to the nearest single decimal percentage (i.e., 10%, …, 100%) to improve clarity for respondents. The worst and best levels for the duration of screening and waiting time until results attributes were established in the panel session. Finally, the preferential order of mode and modality of screening cannot be assumed a priori (e.g., it will differ between persons whether the GP location is preferred over the hospital location). Respondents in the preference survey were therefore asked to rank the modes of screening from least burdensome to most burdensome and they were asked to indicate their most preferred location. An overview of the attributes and their levels is presented in Table 1.

Weighting Attributes Using a Preference Survey

The aim of the weighting step is to identify the relative importance of attributes with attribute weights. These attribute

Table 1 – Included attributes and levels.

Attribute	Attribute name	Worst level	Best level
Continuous attributes	Sensitivity	70%	100%
	Specificity	70%	100%
	Radiation burden	Background + 1.5 mSv	Background
	Duration of screening procedure	45 min	15 min
	Waiting time until results	14 d	1 d
Categorical attributes	Scan type	Levels: "lie in scanner," "sustained breath into device," "give blood"	
	Location	Levels: "nearest hospital," "at your GP's office"	

Note. Sensitivity is the probability of a positive/suspect test result given that the person has lung cancer. Specificity is the probability of a negative test result given that the person does not have lung cancer. The worst and best levels for the continuous attributes were defined on the basis of interviews with clinicians. For the categorical attributes, the preference order was elicited from respondents in separate questions. GP, general practitioner.

weights were elicited in the preference study from the respondents using a swing weighting task [27,37]. From this we obtained a set of seven attribute weights for each respondent (see Table 2). These weights indicate the relative importance of improvements from the worst performance level to the best performance level between attributes [38]. One thousand thirty-four respondents from the Dutch general population completed the preference survey. Fifty-one percent of the respondents were women, the mean age was 58 ± 11 years, and the respondents were distributed evenly between low, medium, and higher levels of education. For more details regarding the attribute selection process, the preference elicitation process, and the characteristics of the population sample, you may refer to the study by Broekhuizen et al. [27].

Measuring the Performance Data of Screening Programs

The aim of the scoring step is to find out how well screening programs perform on the included attributes. To be able to do this, we collected clinical evidence on the screening programs' performances. When this was not available, we made assumptions on the basis of the clinical experience of two of the authors (R.V. and H. G.). For the sensitivity and specificity of LDCT, published pooled estimates from the first round of the Dutch-Belgian Randomized Lung Cancer Screening Trial (NELSON) were used [39]. For the sensitivity and specificity of BA, we included trials from a recent systematic review [32]. For the estimation of the predictive power of BA in terms of the sensitivity and specificity, a prediction model was constructed on the basis of the chemical patterns found in known patients with cancer and known healthy persons. Prediction models constructed in this way are susceptible to "overfitting," an undesirable effect when the model memorizes the persons in the case-control study instead of identifying generalizable patterns and this leads to poor predictive power for new cases outside of the original case-control study. We excluded trials from the review that did not correct for overfitting because their performance estimates (mostly very close to 100%) were unlikely to generalize well [32]. We subsequently pooled the sensitivity and specificity estimates from the eight included clinical trials with a bivariate copula random effects model in which the choice of copula function was made with maximum-likelihood estimation [40]. Performance data for sensitivity and specificity of BB were also obtained from trials cited in a systematic review [35]. As for BA, the estimates for BB were pooled using a bivariate copula random effects model [40]. The radiation burden for LDCT was set to 1.5 mSv, and the radiation burden for BA and BB was set to 0 mSv. For duration of screening we assumed that the participant would need between 15 and 45 minutes for LDCT, between 20 and 30 minutes for BA, and between 10 and 25 minutes for BB. LDCT images have to be evaluated by a radiologist and we assumed that it would take 1 week for the participant to receive the result. Samples for BB are mostly analyzed in separate laboratory facilities and therefore the time until results for BB was set to 2 days. We assumed that the BA results would be available the next day because the BA review authors noted that "time between breath collection and analysis was usually a couple of hours" [32]. We assumed that LDCT screening is performed at the local hospital. Screening based on BA or BB is assumed to be done at a primary care facility. Table 2

Finally, the three screening technologies directly corresponded to the levels for the mode of screening attribute: for LDCT one "goes through a scanner," for BA one "has to exhale into

Table 2 – Preference data.

		Bisection point percentiles		
Attribute name	Weight, mean ± SD	2.5%	50%	97.5%
Sensitivity	0.16 ± 0.14	63.5%	90.0%	100.0%
Specificity	0.13 ± 0.12	64.3%	90.4%	100.0%
Radiation burden	0.13 ± 0.12	0.16 mSv	1.10 mSv	1.5 mSv
Duration of screening procedure	0.10 ± 0.09	42.0 min	26.4 min	15.0 min
Waiting time until results	0.13 ± 0.12	12.8 d	5.2 d	1 d
Mode of screening	0.17 ± 0.14	21.0%	72.5%	100.0%
Location of screening	0.18 ± 0.17	–	–	–

Note. For mode of screening, the bisection point indicates the respondent's indifference between an improvement from the worst level to the bisection point and the improvement from the bisection point to the best level. No bisection point was elicited for the location of screening because it had only two (categorical) levels. The lower and upper levels per attribute are presented in Table 1. More details can be found in Broekhuizen et al. [27].

a device," and for BB one has to "give blood." An overview of all performance inputs is presented in Table 3.

Aggregating into Overall Values

In the aggregation step, a value score is calculated for each screening program. Because performances are measured on different natural scales (percentages, days, etc.) we first need to transform these metrics to a partial value between 0 (no value) and 1 (maximum value). Finally, the value estimate for each included screening policy is obtained by weighting the partial values with the attribute weights.

A partial value function is a mathematical function that maps attribute-specific performances (percentages, days, etc.) onto a partial value scale between 0 and 1. Although partial value functions are commonly assumed to be linear, we considered this to be unrealistic in our case because there are two time-related attributes in our attribute set and people's preferences for time are often nonlinear [41]. Moreover, we considered it unlikely that the difference in preference between the first- and second-ranked screening modes was equal to the difference in preference between the second- and third-ranked screening modes.

To estimate a nonlinear partial value function for the continuous attributes, we elicited from each respondent a bisection point on the performance scale of each attribute (except for location). We excluded 48 respondents who did not complete the bisection point questions. The bisection point indicates the respondent's indifference between an improvement from the worst level to the bisection point and the improvement from the bisection point to the best level. Denoting the indifference point with x^*_{kq} and the worst and best levels with x^-_k and x^+_k, respectively, we obtained the two-piece partial value function v_{kq} of continuous attribute k (k = 1,...,5) for respondent q (q = 1,...,n):

$$v_{kq}\left(x_{ki};x^*_{kq},x^-_k,x^+_k\right)=\begin{cases} 0, & \text{if } x_{ki}\le x^-_k \\ \frac{1}{2}\left(\frac{x_{ki}-x^-_k}{x^*_{kq}-x^-_k}\right), & \text{if } x^-_k<x_{ki}<x^*_{kq} \\ \frac{1}{2}\left(1+\frac{x_{ki}-x^*_{kq}}{x^+_k-x^*_{kq}}\right), & \text{if } x^*_{kq}\le x_{ki}<x^+_k \\ 1, & \text{if } x_{ki}\ge x^+_k \end{cases}. \quad (1a-d)$$

Here, x_{ki} is the performance of screening program i on attribute k. The sixth attribute, mode of screening, is categorical. Therefore, we defined the partial value $v_{6,q}$ for the third-ranked mode for a respondent q as having a value of 0, the second-rank mode as having a value of $x^*_{6,q}$ (the bisection point), and the first-ranked mode as having a value of 1. The seventh attribute (location of screening) is also categorical with two levels, and so the partial value $v_{7,q}$ for the second-ranked location was 0 and that for the first-ranked location was 1.

To aggregate the respondents' partial values for the different attributes into an overall value for each included screening program (denoted as V_{iq}), the partial values were scaled using the attribute weights and then summed:

$$V_{iq}=\sum_{k=1}^{7}w_{kq}v_{kq}. \quad (2)$$

Then, an estimate of the mean overall value for a screening program in the general population (denoted as V_i), was obtained by averaging the individual overall values over all n respondents:

$$V_i=\frac{\sum_{q=1}^{n}V_{iq}}{n}. \quad (3)$$

By sorting the screening programs per respondent from the highest value to the lowest value, a rank order of screening programs per respondent r_{iq} was obtained. Furthermore, we denoted the rank of program i with R_i. The proportion of the population for whom this rank equals y = 1, 2, or 3 was then estimated with the proportion of respondents for whom program i had rank y (Equations 4 and 5):

$$\frac{\sum_{q=1}^{n}1_y(r_{iq})}{n}, \quad (4)$$

where

$$1_y\left(r_{iq}\right)=\begin{cases} 1, & \text{if } r_{iq}=y \\ 0, & \text{otherwise} \end{cases}. \quad (5)$$

Uncertainty Analyses

We investigated the impact of parameter uncertainty in clinical evidence and heterogeneity in attribute weights on the mean overall values of the screening programs [42,43]. Parameter uncertainty was investigated using a probabilistic approach. The uncertain parameter estimators in our model were those for the programs' performance on the sensitivity, specificity, and duration of screening attributes. For LDCT, we modeled the uncertainty in the parameter estimates for sensitivity and specificity by two independent beta distributions according to best practices for modeling proportions [44]. We also assumed that the average screening duration for LDCT was uniformly distributed between 15 and 45 minutes. For both BA and BB, the uncertainty in the estimators for sensitivity and specificity was modeled with the bivariate copula distributions that were fitted for the pooling of study-level estimates [40]. Furthermore, we assumed that the average screening duration was uniformly distributed between 20 and 30 minutes for BA and between 10 and 25 minutes for BB.

Table 3 – Used performance data of the three included screening policies with 95% CIs (for sensitivity and specificity) or ranges (for duration of screening).

Attribute name	LDCT	BA	BB
Sensitivity (%)	92.5 (85.3–97.5)	80.0 (52.8–96.7)	74.8 (66.7–85.2)*
Specificity (%)	98.8 (98.0–98.6)	79.8 (66.7–89.9)	87.9 (77.1–95.7)†
Radiation burden (mSv)	1.5	0	0
Duration of screening (min)	30 (range 15–45)	25 (range 20–30)	17.5 (range 10–25)
Waiting time until results (d)	7	1	2
Mode	Go through scanner	Exhale into device	Give blood
Location	Hospital	GP's office	GP's office

BA, breath analysis; BB, blood biomarkers; CI, confidence interval; GP, general practitioner; LDCT, low-dose computed tomography.

* Pooled with Frank copula with Kendall τ=0.9.

† Pooled with normal copula with Kendall τ=1.0.

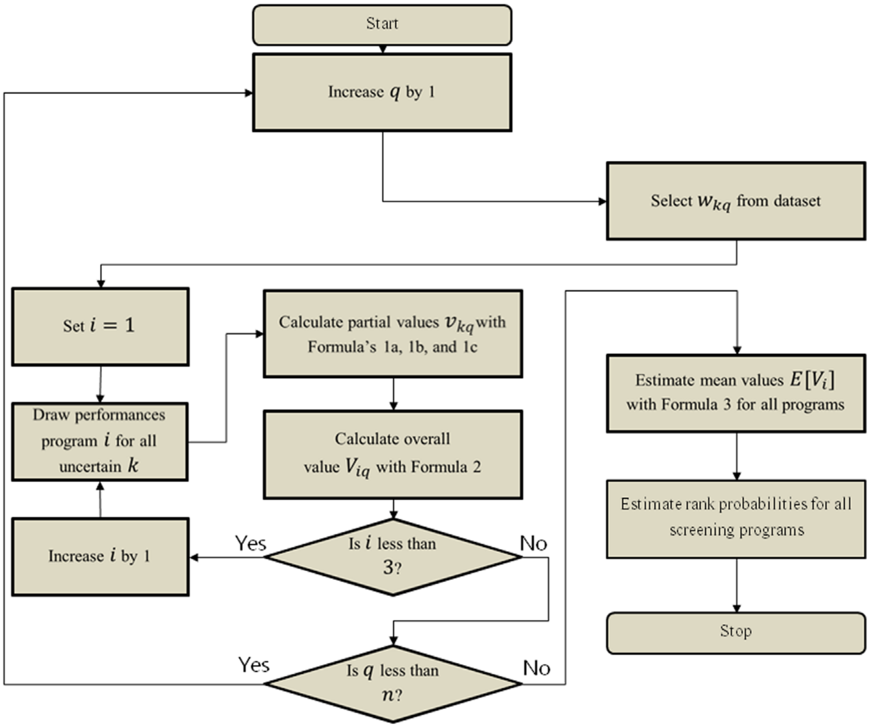
Start
Increase q by 1
Select w_{kq} from dataset
Set $i = 1$
Draw performances program i for all uncertain k
Calculate partial values v_{kq} with Formula's 1a, 1b, and 1c
Calculate overall value V_{iq} with Formula 2
Is i less than 3?
Yes
No
Increase i by 1
Is q less than n?
Yes
No
Estimate mean values $E[V_i]$ with Formula 3 for all programs
Estimate rank probabilities for all screening programs
Stop

Fig. 1 – Modeling flowchart.

Because the overall value of each screening program is a combination of weights and performances, it is hard to analytically find the distribution of V_i given the parameter uncertainty in sensitivity, specificity, and duration. We therefore estimated this value distribution using Monte-Carlo simulations [45]. This meant that we drew one sample from the aforementioned probability distributions of the performance estimators for each respondent and then calculated the partial values of each program for each respondent with Equations 1a to 1c and the overall values using Equation 2. We also calculated the 95% confidence intervals (CIs) of the partial and overall values across the preference sample by taking the empirical 95% CI from the Monte-Carlo simulations. A flowchart of the Monte-Carlo process is shown in Figure 1.

To investigate whether heterogeneity in preferences would influence which of the screening programs had the highest value, we performed the following subgroup analyses. Each scenario consisted of a subgroup from the preference study. First, with cluster analysis the respondents were divided into five subgroups of persons with similar preferences [27]. Each of these preference subgroups is named after the attribute(s) respondents in the subgroup considered most important: radiation-sensitivity (n = 236), waiting time until results (n = 157), location-mode (n = 299), mode-sensitivity (n = 184), and sensitivity-specificity (n = 158). Apart from these preference subgroups, we investigated subgroups in the preference study that can be defined in terms of their eligibility for screening according to three recent screening trials. Each of these trials (NLST, NELSON, and the UK Lung Cancer Screening [UKLS]) had different eligibility criteria based on a person's risk profile. The NLST study included men and women aged between 55 and 74 years who had more than 30 pack-years of smoking [3]. Furthermore, former smokers must have quit within the previous 15 years. A pack-year is defined as 20 cigarettes smoked every day for 1 year. In the preference study, it was not asked how many cigarettes respondents smoked per day but instead it was asked how many years respondents had smoked and if they had quit, how long ago that was. Because the average number of cigarettes smoked in the Netherlands by smokers in the eligible population is 11 [46], we multiplied the total number of years smoked by 11/20 = 0.55 to estimate the respondents' pack-years. Here, we also assume that the daily

Table 4 – Mean partial values (as calculated for the base case with Equation 1) with 95% CIs.

Attribute name	LDCT	BA	BB
Sensitivity	0.63 (0.62–0.64)	0.39 (0.37–0.41)	0.33 (0.32–0.34)
Specificity	0.81 (0.80–0.82)	0.37 (0.36–0.38)	0.52 (0.51–0.53)
Radiation burden	0.00 (0.00–0.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)
Duration of screening	0.17 (0.16–0.18)	0.23 (0.23–0.23)	0.53 (0.51–0.55)
Waiting time until results	0.26 (0.26–0.26)	1.00 (1.00–1.00)	0.45 (0.45–0.45)
Mode	0.65 (0.63–0.67)	0.37 (0.34–0.40)	0.68 (0.65–0.71)
Location	0.40 (0.37–0.43)	0.60 (0.57–0.63)	0.60 (0.57–0.63)

Note. 95% CIs were obtained with Monte-Carlo simulations.
BA, breath analysis; BB, blood biomarkers; CI, confidence interval; LDCT, low-dose computed tomography.

Table 5 – Overview of values and rank proportions.

Mode	Parameter uncertainty	Stochastic uncertainty					
	Mean overall value (CI)	Value percentiles			Percent of respondents for whom modality was ranked …		
		2.5%	50%	97.5%	First	Second	Third
LDCT	0.44 (0.43–0.45)	0.16	0.44	0.74	23.4%	19.7%	56.9%
BA	0.57 (0.56–0.59)	0.26	0.58	0.87	32.7%	45.0%	22.3%
BB	0.58 (0.57–0.59)	0.26	0.60	0.87	43.9%	35.3%	20.8%

Note. The CI for the mean overall values reflects uncertainty on the sample level because of uncertainty in parameter estimates for clinical performance. The results under "Stochastic uncertainty" reflect differences at the respondent level.
BA, breath analysis; BB, blood biomarkers; CI, confidence interval; LDCT, low-dose computed tomography.

number of cigarettes smoked by someone is constant. The NELSON study included men and women aged between 50 and 78 years who smoked either more than 15 cigarettes every day during more than 25 years or more than 10 cigarettes every day during more than 30 years [47]. Furthermore, to be eligible for the NELSON study one needs to be a current smoker or a former smoker who quit within the previous 10 years. Finally, the UKLS study included men and women aged between 50 and 75 years with 5-year lung cancer risk of more than 5% as calculated with the Liverpool Lung Project model [48,49]. Note that the subgroups based on the screening trials were not mutually exclusive; that is, persons with a very high risk of lung cancer are likely to be eligible for multiple trials. For each of the eight subgroups, we reran the model, including parameter uncertainty in estimates for sensitivity, specificity, and duration of screening. In this way we obtained mean value estimates and associated CIs of the screening programs according to each subgroup. By comparing these mean values, we assessed which screening program would be most valuable for different subgroups.

Results

The partial values of the screening programs for sensitivity and specificity reflected the differences in the clinical evidence: LDCT had the highest partial value for both sensitivity and specificity, BA had a higher partial value for sensitivity than did BB, and BB had a higher partial value for specificity than did BA. LDCT had the lowest partial value for the duration of screening and waiting time until results attributes, BB had the highest partial value for the duration of screening attribute, and BA had the highest partial value for the waiting time until results attribute. Both LDCT and BB had similar partial values for the mode of screening attribute (about 0.65), whereas BA had a clearly lower partial value for that attribute (0.37). Because both BA and BB were assumed to take place at the GP's office and both had no radiation burden, the partial values for this attribute were identical. An overview of the partial values is presented in Table 4.

BB had a mean overall value of 0.58 (95% CI 0.57–0.59) and BA had a mean overall value of 0.57 (95% CI 0.56–0.59). Although the mean overall value for BB was slightly higher than that of BA, the absolute difference was small (0.01) and the CIs due to parameter uncertainty overlapped. LDCT, however, clearly had the lowest mean overall value (0.44), and its CI (0.43–0.45) did not overlap with that of either BA or BB. Although the distribution of the overall value is skewed to the left, the median overall values coincide with the mean. In contrast with the CIs around estimates of the mean overall values, which were small, the individual value distributions showed considerable overlap. This is reflected in the rank proportions: BB had the highest value for 43.9% of respondents, whereas this was 32.7% for BA and 23.4% for LDCT. For 56.9% of respondents, LDCT had the last rank. An overview of the overall values and the rank probabilities is presented in Table 5. For most preference subgroups, the overall values were similar to those of the entire sample (Table 6; Fig. 2). For the respondents in the waiting time until results preference subgroup, however, BA had the highest value and its CI did not overlap with those of the other screening programs. In the sensitivity-specificity preference subgroup, LDCT had the highest mean value although its CI overlapped with those of the other programs. The preference for BB according to respondents in the radiation-sensitivity preference subgroup was stronger than that of the entire sample in that the CI for BB no longer overlapped with that of BA. From the sample, 10.3% to 18.1% were eligible for one of the trials (see Tables 7 and 8). The attribute weights as assigned by respondents in these particular subgroups were

Table 6 – Overall values (with 95% CIs) for the three different screening programs according to each of the included subgroups.

Preference subgroup	LDCT	BA	BB
Radiation-sensitivity	0.44 (0.42–0.46)	0.57 (0.55–0.59)	0.63 (0.61–0.65)
Mode-sensitivity	0.47 (0.45–0.49)	0.56 (0.53–0.59)	0.55 (0.52–0.58)
Location-mode	0.42 (0.39–0.44)	0.58 (0.56–0.60)	0.61 (0.58–0.64)
Waiting time until results	0.36 (0.35–0.38)	0.66 (0.64–0.68)	0.58 (0.56–0.60)
Sensitivity-specificity	0.54 (0.52–0.56)	0.50 (0.47–0.52)	0.51 (0.49–0.53)
NLST inclusion criteria	0.47 (0.44–0.50)	0.56 (0.53–0.59)	0.53 (0.51–0.56)
NELSON inclusion criteria	0.47 (0.45–0.50)	0.55 (0.52–0.57)	0.55 (0.52–0.58)
UKLS inclusion criteria	0.51 (0.47–0.54)	0.54 (0.50–0.58)	0.51 (0.47–0.54)

BA, breath analysis; BB, blood biomarkers; CI, confidence interval; LDCT, low-dose computed tomography; NELSON; NLST, National Lung Screening Trial; UKLS, UK Lung Cancer Screening.

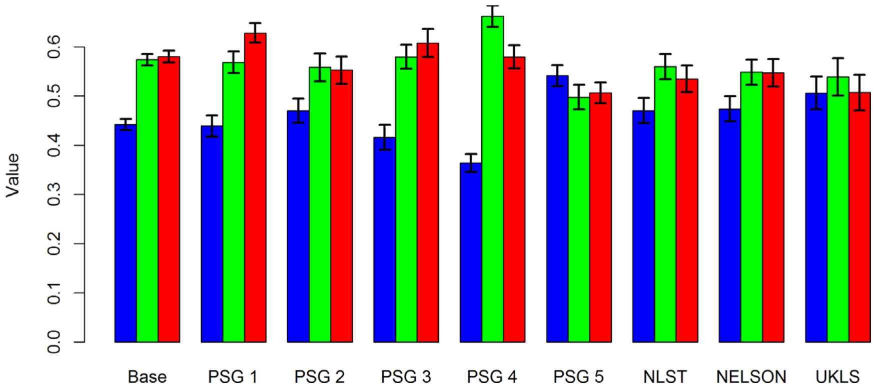

Fig. 2 – Overall mean values for LDCT (blue), breath analysis (green), and blood biomarkers (red) with 95% CI. "Base" denotes the total sample and "PSGx" denotes the *x*th preference subgroup (see Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2018.01.021 and Tables 7 and 8 for an overview). CI, confidence interval; LDCT, low-dose computed tomography; NLST, National Lung Screening Trial; UKLS, UK Lung Cancer Screening.

similar to those of the whole sample. Nevertheless, they seemed to consider the burden of exhaling into a device more severe and giving blood less severe than the sample as a whole [27]. Also, although in the entire sample 60% prefer a GP's office to a hospital, this is only 53% for those eligible for the NLST trial. In the subgroup of respondents who are eligible for the UKLS trial, 48% of respondents prefer the GP's office. With regard to the overall values for these risk subgroups, we see that the overall values for the NLST and NELSON subgroups are similar to those of the entire sample; that is, BA and BB have similar values and LDCT has a significantly lower value. For the respondents who were eligible for the UKLS trial, BA had the highest value although the CIs of all three programs overlapped.

Discussion

In this study, we have assessed the relative value of three lung cancer screening policies using a probabilistic MCDA model. BB had the highest mean overall value but the difference between BB and BA is so small that it is unlikely to have policy relevance. We do see that LDCT had the lowest mean overall value and this is stable given the parameter uncertainty in estimates for clinical performance. Furthermore, LDCT had the lowest mean overall value in most subgroups we investigated. It is important to consider that the present study was based on a public sample, from which only 10% to 18% is currently eligible for screening under guidelines used by recent lung cancer screening trials. If only the preferences of those eligible for screening according to the criteria used in the NELSON or NLST trial are taken into account, it can be seen that the overall values for the screening programs are very similar to those found in the entire sample. For the respondents who are eligible for screening according to the UKLS trial criteria, BA has the highest mean value but the CIs around the mean overall values for all three screening programs overlap. Although our analyses show that the mean values for BA and BB are higher than for LDCT, there is still considerable variation in the value of screening programs for individual respondents as can be seen in the ranking proportions. It seems that most respondents focus on nonclinical attributes, whereas only 15% (namely those in the sensitivity-specificity subgroup) focus on diagnostic accuracy. Unsurprisingly, LDCT had the highest mean value for this subgroup. For the other subgroups and the sample as a whole, LDCT had the lowest mean overall value. This implies that nonclinical attributes of screening play an important role in people's preferences for screening programs, explaining why LDCT has such a low mean value. Nevertheless, the high mean values for BA and BB found in our study may be

Table 7 – Attribute weights per respondent subgroup.

		Attribute weight, mean ± SD						
Subgroup based on	n (%)	Se	Sp	Ra	Du	Wa	Mo	Lo
Preference subgroup preferences								
PSG1	227 (23)	0.21 ± 0.12	0.20 ± 0.11	0.25 ± 0.17	0.09 ± 0.07	0.07 ± 0.05	0.12 ± 0.10	0.07 ± 0.05
PSG2	174 (18)	0.15 ± 0.11	0.10 ± 0.06	0.09 ± 0.06	0.11 ± 0.11	0.11 ± 0.06	0.31 ± 0.20	0.12 ± 0.07
PSG3	284 (29)	0.08 ± 0.06	0.06 ± 0.04	0.10 ± 0.06	0.10 ± 0.08	0.11 ± 0.07	0.18 ± 0.12	0.36 ± 0.18
PSG4	151 (15)	0.09 ± 0.06	0.07 ± 0.04	0.14 ± 0.09	0.13 ± 0.12	0.29 ± 0.18	0.14 ± 0.08	0.14 ± 0.07
PSG5	150 (15)	0.30 ± 0.19	0.26 ± 0.17	0.07 ± 0.05	0.08 ± 0.07	0.11 ± 0.07	0.09 ± 0.05	0.09 ± 0.06
Population according to trial eligibility								
NLST	188 (18)	0.15 ± 0.14	0.12 ± 0.10	0.12 ± 0.10	0.10 ± 0.09	0.14 ± 0.12	0.18 ± 0.16	0.18 ± 0.16
NELSON	183 (18)	0.16 ± 0.15	0.13 ± 0.12	0.11 ± 0.10	0.10 ± 0.09	0.14 ± 0.12	0.17 ± 0.16	0.18 ± 0.15
UKLS	106 (10)	0.15 ± 0.14	0.12 ± 0.11	0.10 ± 0.08	0.10 ± 0.08	0.14 ± 0.13	0.20 ± 0.18	0.18 ± 0.16

Note. The preference subgroup names are based on the two most important attributes in that cluster [27]. The mean weights with SDs are presented per subgroup as well.

Attributes: Se, sensitivity; Sp, specificity; Ra, radiation burden; Du, duration of screening; Wa, waiting time until results; Mo, mode of screening; Lo, location of screening; NELSON; NLST, National Lung Screening Trial; PSGx, preference subgroup number *x* (here 1 is radiation-sensitivity, 2 is mode-sensitivity, 3 is location-mode, 4 is waiting time until results, and 5 is sensitivity-specificity); UKLS, UK Lung Cancer Screening.

Table 8 – Mode and location ranks per respondent subgroup.

	First-ranked mode (% of respondents in subgroup)			First-ranked location (% of respondents in subgroup)
	LDCT	BA	BB	GP
Preference subgroup				
PSG1	16	59	25	57
PSG2	32	34	34	61
PSG3	23	39	38	65
PSG4	28	46	26	56
PSG5	21	50	29	55
Population according to				
NLST	25	38	37	53
NELSON	20	43	37	55
UKLS	28	30	42	48

Note. The preference subgroups names are based on the two most important attributes in that cluster [27].
BA, breath analysis; BB, blood biomarkers; GP, general practitioner; LDCT, low-dose computed tomography; NELSON; NLST, National Lung Screening Trial; PSGx, preference subgroup number *x* (here 1 is radiation-sensitivity, 2 is mode-sensitivity, 3 is location-mode, 4 is waiting time until results, and 5 is sensitivity-specificity); UKLS, UK Lung Cancer Screening.

biased by the low quality level of evidence. No prospective clinical trials have been done comparing BA or BB with LDCT, and the diagnostic accuracy estimates we use are for clinical tumors, not for preclinical tumors which would be more relevant in a screening setting. In contrast, multiple large prospective screening trials have been performed for LDCT, which provide confidence and a solid and positive argument for its use in screening. It would therefore be better to consider our results in light of the question of how a screening program based on LDCT can be further improved and implemented.

Our results can support policymakers in defining strategies that can help the screening program reach as many people as possible. Two practical concerns that reduce the value of LDCT are waiting time until results and radiation burden. An improvement strategy may thus be to try and reduce waiting time until results are communicated to people, for example, by evaluating CT images using machine learning algorithms [50]. Next, the fact that the respondents in our sample seemed quite worried about the radiation burden posed by LDCT (mean weight of 0.13) is interesting because the actual incremental risk due to the radiation induced by LDCT is likely to be small. A strategy for addressing these concerns may be targeted information campaigns that inform and shift people's views. If machine learning algorithms could reduce the waiting time until results to 1 day and if the information campaign would reduce the concerns about the radiation burden, our model shows that the overall value for LDCT would be close or equal to BA and BB. This would be beneficial because it improves people's satisfaction with the screening program, subsequently leading to more people attending first or follow-up screening rounds. This increased attendance would increase the cost effectiveness of the screening program as a whole.

The main strength of our study is that we had a large representative sample from the Dutch population in which we could study the value of screening programs for subgroups on the basis of preferences or eligibility for trials. Second, by eliciting bisection points we improved upon the commonly used but unrealistic assumption of linear partial value functions in MCDA for health care. Third, we provided an evidence-based assessment of the perceived value of lung cancer screening policies by explicitly including clinical evidence and results from a stated preference study. As stated earlier, the main limitation of our study is that the quality of evidence for BA and BB was low. This precludes us from making strong statements about the value of these modalities in screening practice. Moreover, because we did not know the pack-years of the respondents but only the years of smoking, we had to make assumptions regarding people's smoking habits on the basis of publicly available data. Finally, we have taken the perspective of the potential participant who is deciding which screening policy to attend for the first time. We have not taken into account the long-term consequences of screening or perceived values relevant for subsequent screening rounds.

Conclusions

We have evaluated the values of three different screening policies by combining data from a large public preference sample and clinical evidence. Although the evidence base for lung cancer screening based on BA or BBs is still weak, these screening modalities seem promising because of their practical advantages such as application in a primary care setting. To improve the value of LDCT-based screening for the eligible population it is important to take nonclinical attributes of screening into account.

Source of financial support: No funding was received for the preparation of this article.

Supplemental Materials

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.jval.2018.01.021. The patient preference data set is available from the corresponding author on reasonable request.

REFERENCES

[1] Optican RJ, Chiles C. Implementing lung cancer screening in the real world: opportunity, challenges and solutions. Transl Lung Cancer Res 2015;4:353–64.

[2] Kauczor H-U, Bonomo L, Gaga M, et al. ESR/ERS white paper on lung cancer screening. Eur Respir J 2015;46:28–39.

[3] The National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med 2011;365:395–409.

[4] Molina JR, Yang P, Cassivi SD, et al. Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship. Mayo Clin Proc 2008;83:584–94.

[5] Gulati S, Mulshine JL. Lung cancer screening guidelines: common ground and differences. Transl Lung Cancer Res 2014;3:131–8.
[6] McMahon PM, Meza R, Plevritis SK, et al. Comparing benefits from many possible computed tomography lung cancer screening programs: extrapolating from the National Lung Screening Trial using comparative modeling. PLoS One 2014;9:e99978.
[7] de Koning HJ, Meza R, Plevritis SK, et al. Benefits and harms of computed tomography lung cancer screening strategies: a comparative modeling study for the U.S. Preventive Services Task Force. Ann Intern Med 2014;160:311–20.
[8] Adiguzel Y, Kulah H. Breath sensors for lung cancer diagnosis. Biosens Bioelectron 2015;65:121–38.
[9] Xiang D, Zhang B, Doll D, et al. Lung cancer screening: from imaging to biomarker. Biomark Res 2013;1:4.
[10] Chan HP, Lewis C, Thomas PS. Exhaled breath analysis: novel approach for early detection of lung cancer. Lung Cancer 2009;63:164–8.
[11] Ford ME, Havstad SL, Flickinger L, Johnson CC. Examining the effects of false positive lung cancer screening results on subsequent lung cancer screening adherence. Cancer Epidemiol Biomarkers Prev 2003;12:28–33.
[12] Montes U, Seijo LM, Campo A, et al. Factors determining early adherence to a lung cancer screening protocol. Eur Respir J 2007;30:532–7.
[13] Mansfield C, Tangka FKL, Ekwueme DU, et al. Stated preference for cancer screening: a systematic review of the literature, 1990–2013. Prev Chronic Dis 2016;13:E27.
[14] Wildstein KA, Faustini Y, Yip R, et al. Longitudinal predictors of adherence to annual follow-up in a lung cancer screening programme. J Med Screen 2011;18:154–9.
[15] Weernink MGM, Janus SIM, van Til JA, et al. A systematic review to identify the use of preference elicitation methods in healthcare decision making. Pharmaceut Med 2014;28:175–85.
[16] Brett Hauber A, Fairchild AO, Reed Johnson F. Quantifying benefit-risk preferences for medical interventions: an overview of a growing empirical literature. Appl Health Econ Health Policy 2013;11:319–29.
[17] Tony M, Wagner M, Khoury H, et al. Bridging health technology assessment (HTA) with multicriteria decision analyses (MCDA): field testing of the EVIDEM framework for coverage decisions by a public payer in Canada. BMC Health Serv Res 2011;11:329.
[18] Phillips KA, Van Bebber S, Marshall D, et al. A review of studies examining stated preferences for cancer screening. Prev Chronic Dis 2006;3:A75.
[19] Wortley S, Wong G, Kieu A, Howard K. Assessing stated preferences for colorectal cancer screening: a critical systematic review of discrete choice experiments. Patient 2014;7:271–82.
[20] Hummel JM, Steuten LGM, Groothuis-Oudshoorn CJM, et al. Preferences for colorectal cancer screening techniques and intention to attend: a multi-criteria decision analysis. Appl Health Econ Health Policy 2013;11:499–507.
[21] Groothuis-Oudshoorn CGM, Fermont J, van Til JA, et al. Public stated preferences and predicted uptake for genome-based colorectal cancer screening. BMC Med Inform Decis Mak 2014;14:18.
[22] Ghanouni A, Smith S, Halligan S, et al. Public preferences for colorectal cancer screening tests: a review of conjoint analysis studies. Expert Rev Med Devices 2013;10:489–99.
[23] Marshall D, McGregor E, Currie G. Measuring preferences for colorectal cancer screening: What are the implications for moving forward? Patient 2010;3:79–89.
[24] de Bekker-Grob EW, Rose JM, Donkers B, et al. Men's preferences for prostate cancer screening: a discrete choice experiment. Br J Cancer 2013;108:533–41.
[25] Howard K, Salkeld GP, Patel MI, et al. Men's preferences and trade-offs for prostate cancer screening: a discrete choice experiment. Health Expect 2015;18:3123–35.
[26] Kaltoft M, Turner R, Nielsen J, et al. Addressing preference heterogeneity in public policy by combining cluster analysis and multi-criteria decision analysis. Health Econ Rev 2014;5:1–11.
[27] Broekhuizen H, Groothuis-Oudshoorn C, Vliegenthart R, et al. Public preferences for lung cancer screening policies. Value Health 2017;20:961–8.
[28] Dodgson J, Spackman M, Pearman A, Phillips L. Multi-Criteria Analysis: A Manual. London: Department for Communities and Local Government, 2009.
[29] Marsh K, IJzerman M, Thokala P, et al. Multiple criteria decision analysis for health care decision making—emerging good practices: report 2 of the ISPOR MCDA Emerging Good Practices Task Force. Value Health 2016;19:125–37.
[30] Bana E, Costa CA, De Corte J-M, Vansnick J-C. MACBETH. Int J Inf Technol Decis Mak 2012;11:359–87.
[31] Xu DM, Gietema H, de Koning H, et al. Nodule management protocol of the NELSON randomised lung cancer screening trial. Lung Cancer 2006;54:177–84.
[32] Krilaviciute A, Heiss J, Leja M, et al. Detection of cancer through exhaled breath: a systematic review. Oncotarget 2015;6:38643–57.
[33] Sozzi G, Boeri M. Potential biomarkers for lung cancer screening. Transl Lung Cancer Res 2014;3:139–48.
[34] Zhang Z, Ramnath N, Nagrath S. Current status of CTCs as liquid biopsy in lung cancer and future directions. Front Oncol 2015;5:1–11.
[35] Hamilton G, Rath B. Detection of circulating tumor cells in non-small cell lung cancer. J Thorac Dis 2016;8:1024–8.
[36] Eleveld H. Ionising Radiation Exposure in the Netherlands. Bilthoven: The Netherlands: National Institute for Public Health and the Environment, 2003.
[37] Thokala P, Devlin N, Marsh K, et al. Multiple criteria decision analysis for health care decision making—an introduction: report 1 of the ISPOR MCDA Emerging Good Practices Task Force. Value Health 2015;19:1–13.
[38] Belton V, Stewart TJ. Multiple Criteria Decision Analysis: An Integrated Approach (2nd ed.). Dordrecht, The Netherlands: Kluwer Academic Publishers, 2002.
[39] Horeweg N, Scholten ET, de Jong PA, et al. Detection of lung cancer through low-dose CT screening (NELSON): a prespecified analysis of screening test performance and interval cancers. Lancet Oncol 2014;15:1341–50.
[40] Nikoloulopoulos AK. A mixed effect model for bivariate meta-analysis of diagnostic test accuracy studies using a copula representation of the random effects distribution. Stat Med 2015;34:3842–65.
[41] Andreoni J, Kuhn M, Sprenger C. Measuring time preferences: a comparison of experimental methods. J Econ Behav Organ 2015;116:451–64.
[42] Broekhuizen H, Groothuis-Oudshoorn CGM, van Til JA, et al. A review and classification of approaches for dealing with uncertainty in multi-criteria decision analysis for healthcare decisions. Pharmacoeconomics 2015;33:445–55.
[43] Groothuis-Oudshoorn C, Broekhuizen H, van Til J. Dealing with uncertainty in the analysis and reporting of MCDA. In: Marsh K, Goetghebeur M, Thokala P, Baltussen R, eds., Multi-Criteria Decision Analysis to Support Healthcare Decisions. (1st ed.). Cham, Switzerland: Springer, 2017:65–83.
[44] Briggs AH, Weinstein MC, Fenwick EAL, et al. Model parameter estimation and uncertainty: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-6. Value Health 2012;15:835–42.
[45] Broekhuizen H, Groothuis-Oudshoorn CGM, Hauber AB, et al. Estimating the value of medical treatments to patients using probabilistic multi criteria decision analysis. BMC Med Inform Decis Mak 2015;15:1–10.
[46] Statistics Netherlands. Health survey data [Data van gezondheidsenquête]. 2016. Available from: http://statline.cbs.nl/Statweb/publication. [Accessed October 9, 2018].
[47] Oudkerk M, Heuvelmans MA. Screening for lung cancer by imaging: the NELSON study. JBR-BTR 2013;96:163–6.
[48] Baldwin DR, Duffy SW, Wald NJ, 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 2011;66:308–13.
[49] Cassidy A, Myles JP, van Tongeren M, et al. The LLP risk model: an individual risk prediction model for lung cancer. Br J Cancer 2008;98:270–6.
[50] Demir Ö, Yılmaz Çamurcu A. Computer-aided detection of lung nodules using outer surface features. Biomed Mater Eng 2015;26:S1213–22.