BMJ Open Respiratory Research

'Reduced' HUNT model outperforms NLST and NELSON study criteria in predicting lung cancer in the Danish screening trial

Oluf Dimitri Røe ,[1,2] Maria Markaki,[3] Ioannis Tsamardinos,[3,4] Vincenzo Lagani,[5,6] Olav Toai Duc Nguyen,[1,2] Jesper Holst Pedersen,[7] Zaigham Saghir,[8] Haseem Gary Ashraf[8,9]

For numbered affiliations see end of article.

Correspondence to
Dr Oluf Dimitri Røe;
olufdroe@yahoo.no

ABSTRACT

Hypothesis We hypothesise that the validated HUNT Lung Cancer Risk Model would perform better than the NLST (USA) and the NELSON (Dutch-Belgian) criteria in the Danish Lung Cancer Screening Trial (DLCST).

Methods The DLCST measured only five out of the seven variables included in validated HUNT Lung Cancer Model. Therefore a 'Reduced' model was retrained in the Norwegian HUNT2-cohort using the same statistical methodology as in the original HUNT model but based only on age, pack years, smoking intensity, quit time and body mass index (BMI), adjusted for sex. The model was applied on the DLCST-cohort and contrasted against the NLST and NELSON criteria.

Results Among the 4051 smokers in the DLCST with 10 years follow-up, median age was 57.6, BMI 24.75, pack years 33.8, cigarettes per day 20 and most were current smokers. For the same number of individuals selected for screening, the performance of the 'Reduced' HUNT was increased in all metrics compared with both the NLST and the NELSON criteria. In addition, to achieve the same sensitivity, one would need to screen fewer people by the 'Reduced' HUNT model versus using either the NLST or the NELSON criteria (709 vs 918, p=1.02e-11 and 1317 vs 1668, p=2.2e-16, respectively).

Conclusions The 'Reduced' HUNT model is superior in predicting lung cancer to both the NLST and NELSON criteria in a cost-effective way. This study supports the use of the HUNT Lung Cancer Model for selection based on risk ranking rather than age, pack year and quit time cut-off values. When we know how to rank personal risk, it will be up to the medical community and lawmakers to decide which risk threshold will be set for screening.

Key messages

► How to improve the selection of candidates for lung cancer screening.
► Data-driven, multivariable predictive models, when trained with state-of-the-art methods in statistics and machine learning, seem to outperform existing criteria sets based on age, pack year and quit time cut-off values. This study supports the use of the HUNT Lung Cancer Model for selection based on risk ranking rather than NLST and NELSON criteria.
► The study is the first to compare a prediction model performance with the NELSON study criteria and shows that selection for lung cancer screening can be significantly improved.

INTRODUCTION

Lung cancer is the leading cause of cancer mortality worldwide,[1] and early diagnosis is paramount for increasing survival. Currently two studies have shown survival benefit of lung cancer screening; the National Lung Screening Trial (NLST) and the Dutch-Belgian randomised lung cancer screening trial NELSON study (abstract only). The

NLST was the largest prospective trial showing that low-dose high-resolution computed axial tomography (CT) scanning versus X-ray of heavy smokers (>30 pack years, <15 years quit time) ages 55–74 at inclusion time and at 6 years of follow-up reduced lung cancer mortality by 20%.[2] At the World Congress of Lung Cancer 2018, the NELSON results revealed that CT screening showed a 26% reduction in lung cancer deaths at 10 years of study follow-up.[3] However, an estimated 26.7% of those who develop lung cancer in a general US population fulfil the NLST inclusion criteria for CT screening, and for the NELSON such an estimation is not available.[4 5]

In a current European Union position statement recently published in the *Lancet Oncology*, risk stratification is one of the keys to ensure the successful implementation of future low-dose CT screening programmes in Europe.[6] Unfortunately, there is no international consensus on which criteria or models to use for the optimal selection for lung cancer screening.

Several multivariable risk prediction models have been proposed to improve the selection of individuals for lung cancer screening.[7] In addition to NLST's pack years, quit-time, and age, these models also have considered other potential risk factors, such as history of respiratory diseases, exposure to occupational dust (asbestos, coal, silica), socioeconomic status, body mass index (BMI), history of cancer, race, education, forced expiratory volume and biochemical parameters (eg, carcinoembryonic antigen, alpha-fetoprotein, and C reactive protein).[8–10] However, these models and corresponding studies also have a variety of potential issues such as age cut-offs, inclusion of mainly heavy smokers, restricted and/or empirical inclusion of predictors and list-wise exclusion of cases with missing data, all of which call into question the transferability of these models to clinical practice.[11]

To address some of the issues above, we have recently developed a validated risk-assessment model, the HUNT Lung Cancer Model, analysing data from 58 343 ever and never smokers of all ages.[11] The model includes seven predictors automatically selected based on data-driven techniques through backwards feature selection from an initial pool of 36 candidate variables. The HUNT Lung Cancer Model has been externally validated in an independent, prospective dataset containing 45 341 ever-smokers with a concordance index (C-index) 0.879 (95% CI 0.866 to 0.891) and with area under the receiver-operating-characteristic curve 0.87 (95% CI 0.85 to 0.89) for the risk of a lung cancer diagnosis within 6 years. Using a 1.75% threshold for the risk to diagnose the disease within 16 years, sensitivity was 81.9% and specificity 78.3%. According to the model, by screening 22% of ever-smokers one could identify 81.85% of all lung cancers within 6 years and was competitive with the NLST criteria for the same number of screenings.[11] The model is valid for either former or current daily smokers of any smoking burden and quit time, and for all of the adult population (>20 years old). Another advantage of the model is that is does not require knowledge of the history of chronic obstructive lung disease (COPD), the history of lung X-rays, the family or personal cancer history, or the educational history that are often hard to obtain in a reliable fashion.

A competitive methodology to the above model, are the NLST and NELSON criteria. These sets of criteria lead to binary decisions: screen or not screen. In contrast, the HUNT Lung Cancer Model ranks individuals according to the risk of developing lung cancer. To apply it in a clinical setting, one would need to determine a threshold of risk above which to screen individuals. This threshold should be determined considering several public heath factors, such as the screening capacity, cost of screening, and increased risk to the individual due to screening.

The goal of this study is to provide a comparative evaluation of the HUNT Lung Cancer model against the NLST and NELSON criteria. To this end, we have selected the data from the Danish Lung Cancer Screening Trial (DLCST) as it is a large scale, completed screening study, with a long-follow up time. Moreover, the DLCST data contain five out of the seven variables required by the HUNT Lung Cancer Model, unlike other datasets (eg, the NLST data). To cope with the issue of the two missing variables in DLCST, we have retrained (re-fit) the HUNT Lung Cancer Model on the original HUNT2 data using only the commonly measured five variables. The resulting model is called the 'Reduced' HUNT model.

METHODS
DLCST cohort
The DLCST was conducted in Denmark, a prospective lung cancer screening trial randomising participants to a baseline and four annual CT scans versus no follow-up. Between the 1 November 2004 and 31 March 2006, a total of 4104 participants (mean age 58 years; 45% women) were enrolled, starting with an initial (baseline) screening and followed by four annual screening rounds and follow-up until 2015.[12] All cancers diagnosed were histologically or cytologically verified. Participants had to be smokers or former smokers with at least 20 pack years, less than 10 years quit time and age 50–70. The participants also had to be able to walk 36 stair-steps up without stopping, and have lung function (forced expiratory volume in one second) >30% of predicted. The exclusion criteria were; body-weight >130 kg, former lung, breast or kidney cancer or malignant melanoma, >5 years after treatment for other cancers, symptoms of lung cancer (haemoptysis, chest pain, weight loss >6 kg, dyspnoea at rest), CT scan within last year or treatment for tuberculosis less than 2 years ago.

Developing the 'Reduced' HUNT model based on the HUNT Lung Cancer Model methodology
The original HUNT Lung Cancer Model was developed based on the Nord-Trøndelag Health Study 2 (HUNT2), which is a collaboration between HUNT Research Centre (Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology), Nord-Trøndelag County Council, Central Norway Health Authority, and the Norwegian Institute of Public Health.

In short, from 1995 to 1997, HUNT2 invited 93 898 residents of Nord-Trøndelag County in Norway, aged 20 years or more, to participate in a health survey, and ≈70% (n=65 237) responded.[13] The data were collected through questionnaires on demographic characteristics, medical history and lifestyle (199 variables). In 2012, our group was granted access to analyse the HUNT2 data to identify lung cancer cases and establish the HUNT2 discovery dataset. This dataset was linked with the national 11-digit personal identification number of each participant to the Norwegian Cancer and Death Cause Registry. The resulting risk prediction model, called HUNT Lung Cancer Model included seven variables: age, pack years, smoking intensity, years since smoking cessation, BMI,

'daily cough in periods of the year' and 'hours of indoors smoke exposure'.[11]

The ideal scenario would be to test the HUNT Lung Cancer Model with all seven variables in the Danish cohort. However, two variables were not recorded in the DLCST, namely 'daily cough in periods of the year' and 'hours of indoor smoke exposure'. We therefore re-trained a new model to predict lung cancer using the remaining five variables; age, pack years, BMI (height and weight were given and used for BMI estimation), quit years, cigarettes per day, adjusted for sex. This 'Reduced' HUNT model (named HUNT model hereafter) was trained in the HUNT2 subcohort of 12 091 ever-smokers aged 50–70 (online supplementary table S1), which is the DLCST inclusion age, to ensure applicability to the Danish cohort.[1] In this HUNT subcohort, 227 lung cancers were diagnosed within 10 years follow-up. The same statistical methodology as described in Markaki *et al*[11] was applied, excluding the feature selection step; the five variables measured in DLCST were by default included in the model. In more detail, the original variables were non-linearly transformed whenever necessary, as in the original paper; hence pack years, quit-time and BMI were logarithmically transformed. Missing values were imputed using multiple imputation with predictive mean matching (R package mice), resulting in 30 complete datasets. For each of them, 200 bootstrap datasets were generated for the internal validation of the model using R package rms.[14] Discrimination power measured by the C-index and calibration were assessed as performance metrics. In calibration plots, the Hosmer-Lemeshow test was used to denote goodness of fit between the predicted and observed individual risks. To analyse the model with the clinical criteria on equal grounds, we set a risk threshold that defines equal number of screenings to the ones suggested by NLST or NELSON and compare the numbers of false positives and false negatives with the χ^2 test (p<0.05 deemed significant). The χ^2 test was also used for comparing percentage of screenings according to the HUNT model to achieve the same sensitivity as the NLST and NELSON (p<0.05 deemed significant).

To identify a high-risk group, we consider as a high-risk individual anybody with a risk score in the top 16 quantile, as proposed by Royston and Altman and used in the original HUNT model study.[11 15]

NLST and NELSON study criteria

The criteria for inclusion in the NLST are the following: age 55–74, >30 pack years and <15 years quit time. For the NELSON study, the respective criteria are: age 50–75, >15 cigarettes a day >25 years, or >10 cigarettes a day >30 years, 10 or less years quit time. The two sets of criteria were enforced on the DLCST to characterise individuals as high-risk and predict that they would develop lung cancer.

RESULTS

DLCST cohort descriptive statistics

The DLCST included 4104 individuals (2052 screened, 2052 non-screened) where 4051 (98.7%) individuals had registered all the five variables required for the HUNT model and 149 of 153 (97.3%) lung cancers developed in this group. The univariate distributions of each of the five measured variables, as well as sex do not significantly differ between the screened and non-screened groups (table 1).

Performance of the HUNT model on the DLCST cohort

Internal validation in the HUNT2 subset produced a C-index=0.783. The C-index metric equals the probability that the model assigns a higher risk to the person that experiences the event first (lung cancer diagnosis) between a pair of randomly chosen individuals. External validation on the Danish cohort produced a lower C-index (total group 0.663, non-screened 0.709, screened 0.638). The calibration was reliable for the HUNT2 subset (p=0.452) and marginal for the Danish cohort (p=0.0681) (online supplementary figure 1). This finding is explained by the fact that the HUNT2 and DLCST have significantly different distributions of risk factors (table 2) due to the fact that the DLCST is a selected, heavy-smoker population.

Setting as threshold the top 16 quantile risk score of the HUNT in the DLCST would have selected for

Table 1 The screened and non-screened groups in the DLCST (n=4051) regarding the five variables in the 'Reduced' HUNT model; sex, age, pack years, BMI, quit time and cigarettes per day (iqd: interquartile distribution). Variables in the screened and non-screened population of DLCST

Variables	All	Screened	Non-screened	P value
Sex, males (%)	54.70	56.03	54.63	0.3737
Age median (iqd)	57.61 (53.72–61.21)	57.73 (53.66–61.23)	57.45 (53.79–61.20)	0.7836
Pack years median (iqd)	33.75 (27.00–42.00)	34.00 (27.00–42.22)	33.00 (26.25–42.00)	0.2402
BMI median (iqd)	24.75 (22.65–27.27)	24.74 (22.64–27.38)	24.76 (22.66–27.17)	0.7336
Quit time, years median (iqd)	0 (0–0)	0 (0–0)	0 (0–0)	0.1855
Cigarettes per day median (iqd)	20 (15–20)	20 (15–20)	20 (15–20)	0.4665

BMI, body mass index; DLCST, Danish Lung Cancer Screening Trial.

Table 2 Modified Cox-regression model ('Reduced' HUNT model) of cancer risk for ever smokers in HUNT2, restricted to age 50–70 as in the Danish cohort, with no previous cancer, no cancer at inclusion and 10 years follow-up (n=12 091). Body mass index, pack years and smoking quit time had a non-linear association with lung cancer, and these variables were logarithmically transformed

Variable	P value	Beta coefficient (95% CI)	SE
Age	<0.0001	0.0682 (0.0458 to 0.0905)	0.0114
Sex	0.7514	0.0479 (−0.2304 to 0.3262)	0.1420
Body mass index (log)	0.0017	−1.4879 (−2.4409 to 0.5350)	0.4862
Smoking intensity	0.0469	−0.0331 (−0.0623 to 0.0039)	0.0149
Pack years (log)	<0.0001	1.2066 (0.8150 to 1.5982)	0.1998
Smoking quit time (log)	0.0011	−0.2667 (−0.4235 to 0.1099)	0.0800

screening 148 out of the 149 individuals that developed lung cancer (sensitivity 99.3%, specificity 3.31%, positive predictive value (PPV) 3.77%, negative predictive value (NPV) 99.23%, figure 1, table 3).

Comparison of the HUNT model against the NLST criteria

According to the NLST criteria less than half of the Danish cohort 1870/4051 (46.2%) would be considered eligible for screening. Among those selected, 104/149 cases would have been identified for screening (sensitivity 69.80%, p=1.54e-14, in favour of HUNT, figure 1, table 3).

As mentioned, the NLST criteria provide binary decisions (screen, not screen), while the HUNT model ranks individuals according to predicted risk. To compare the model with the clinical criteria on equal grounds, we set a risk threshold that defines equal number of screenings to the ones suggested by NLST. Hence, for the same number of individuals screened (n=918) suggested by

NLST in the non-screened DLCST cohort (n=1 999), the HUNT model showed increased predictive performance in all metrics: sensitivity, specificity, PPV and NPV (table 4). In addition, for a risk threshold of the HUNT model so that the sensitivity achieved (percentage of detected cases out of all cases) equals the sensitivity of the NLST criteria, the number of suggested screenings by the HUNT model is significantly lower: 709 screenings for the HUNT model versus 918 screenings for the NLST (p=1.02e-11, figure 2).

Comparison of the HUNT model against the NELSON criteria

The NELSON study age criterion was 50–75 but had two sets of smoking criteria; >15 cigarettes a day >25 years, ≤10 years quit time (here called NELSON1), or >10 cigarettes a day >30 years, ≤10 years quit time (here called NELSON2). By using the NELSON1 in the whole DLCST 2360/4051 (58.25%) people were selected for screening resulting in 109/149 (sensitivity 73.15%) cancers being
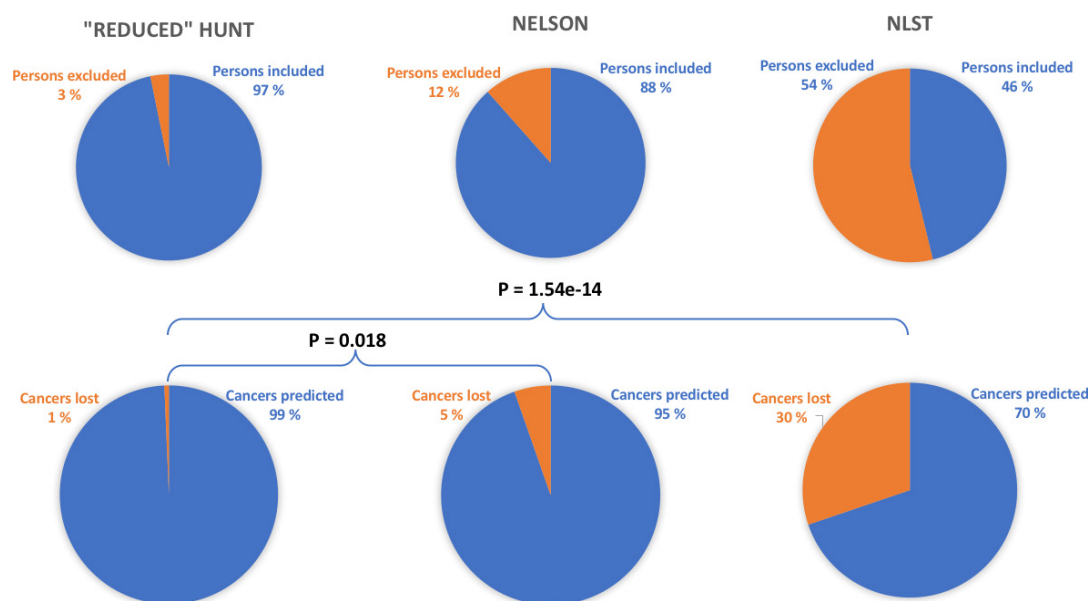

Figure 1 Cartoon of Danish Lung Cancer Screening Trial, screened and non-screened population in total (n=4051, 149 diagnosed cancers). Sensitivity (upper) and cancers 'lost' (lower) using the 'Reduced' HUNT model, the NELSON and the NLST criteria (p values for comparing proportions were by χ^2 test, p<0.05 deemed significant).

Table 3 Performance of the 'Reduced' HUNT model versus NLST and NELSON criteria based on the whole cohort (n=4051). The threshold used in the model to decide screening for lung cancer is the 16% quantile of risk of events in the HUNT cohort with complete data. sensitivity is significantly higher for the HUNT model versus the NLST (p=1.54e-14) and versus the NELSON (p=0.018), but the specificity is lower (figure 1)

	LC (n)	Without LC (n)	Total (n)	Predictive value
Population	149	3902	4051	
'Reduced' HUNT*				
Criteria positive	148 TP (3.77%)	3773 FP (96.23%)	3921	PPV 3.77%
Criteria negative	1 FN (0.77%)	129 TN (99.23%)	130	**NPV 99.23%**
Sensitivity	**99.33%**			
Specificity		3.31%		
NLST†				
Criteria positive	104 TP (5.56%)	1766 FP (94.44%)	1870‡	PPV 5.56%
Criteria negative	45 FN (2.06%)	2136 TN (97.94%)	2181	NPV 97.94%
Sensitivity	**69.80%**			
Specificity		54.74%		
NELSON§				
Criteria positive	141 TP (4.08%)	3449 FP (96.07%)	3590	PPV 4.08 %
Criteria negative	8 FN (1.71%)	461 TN (98.29%)	469	NPV 98.29%
Sensitivity	**94.63%**			
Specificity		11.79%		
NELSON2¶				
Criteria positive	141 TP (4.20%)	3209 FP (95.79%)	3350	PPV 4.20 %
Criteria negative	8 FN (1.14%)	693 TN (98.29%)	701	NPV 98.85%
Sensitivity	**94.63%**			
Specificity		17.76%		

*'Reduced' HUNT model criteria.
†NLST criteria (: >30 pack years of smoking, <15 years since quitting and ages between 55 and 74 years).
‡Total criteria positive selected by the 'Reduced' HUNT model includes those picked by the NLST.
§NELSON criteria age 50–75, >15 cigarettes a day >25 years, or >10 cigarettes a day >30 years, both with 10 or less years quit time.
¶NELSON2 criteria >10 cigarettes a day >30 years, both with 10 or less years quit time.
FN, false negative; FP, false positive; NPV, negative predictive value; PPV, positive predictive value; TN, true negative; TP, true positive.

predicted. The NELSON2 criteria would select 3350/4051 (82.69%) people for screening detecting 141/149 (sensitivity 94.73%) cancers. The NELSON criteria, which is the union of the NELSON1 and NELSON2 criteria, would select 3590/4051 (88.62%) people for screening predicting 141/149 cancers (sensitivity 94.73%, p=0.018, in favour of HUNT) (table 3, figure 1)).

To compare the NELSON criteria to the HUNT model on equal grounds we follow the same methodology described above. When the risk-threshold is set so that the two models suggest the same number of screenings (n=1 668), the HUNT model outperforms the NELSON criteria in all metrics, namely sensitivity, specificity, PPV and NPV (table 5). In addition, for a risk threshold of the HUNT model so that the sensitivity achieved (percentage of detected cases out of all cases) equals the sensitivity of the NELSON criteria, the HUNT model requires the screening of a significantly smaller number of individuals, namely 1317 versus 1668 for the NELSON (p=2.2e-16, figure 2).

DISCUSSION AND CONCLUSIONS

The optimal selection of a high-risk population for lung cancer screening is still an unsolved issue. For example, simple age and smoking criteria, like the NLST, would fail to screen two-thirds of smokers that develop lung cancer.[3]

In this study, we used the validated 7-variable HUNT Lung Cancer Model to develop and apply a 5-variable 'Reduced' HUNT model, based on the five available variables in the Danish screening trial. This downgraded-to-5-variables, reduced model outperforms cut-off criteria such as the NLST and the NELSON. In addition, it outperforms these criteria on the high-risk subpopulations for which they were developed.

Comparison of the 'Reduced' HUNT model with the NLST and NELSON

The DLCST cohort was a high-risk cohort by itself and improving the selection further represented a difficult

Table 4 Direct comparison of the 'Reduced' HUNT model against NLST criteria, for the same number of individuals as selected by the NLST (n=918), on the control group of DLCST (n=1999). The HUNT model shows increased predictive performance in all metrics (sensitivity, specificity, PPV, NPV)

	LC (N)	Without LC (N)	Total (N)	Predictive value
Population	52	1 947	1999	
'Reduced' HUNT				
Criteria positive	40 TP (4.36%)	878 FP (95.64%)	918†	**PPV 4.36%**
Criteria negative	12 FN (1.11%)	1069 TN (98.89%)	1081	**NPV 98.89%**
Sensitivity	**76.92%**			
Specificity		**54.90%**		
NLST‡				
Criteria positive	35 TP (3.81%)	883 FP (96.19%)	918*	PPV 3.81%
Criteria negative	17 FN 1.57%	1064 TN 98.43%	1081	NPV 98.43%
Sensitivity	67.31%			
Specificity		54.65%		

'Reduced' HUNT model.
*'Reduced' HUNT model criteria.
†The number of those picked by the NLST, but with the top risk score by the 'Reduced' HUNT model.
‡NLST criteria (: >30 pack years of smoking, <15 years since quitting and ages between 55 and 74 years).
FN, false negative; FP, false positive; NPV, negative predictive value; PPV, positive predictive value; TN, true negative; TP, true positive.

task. Despite of the participants' relative young age (median 57 years), they had high pack year counts (median 33.75) and were mostly current smokers that are known to have a higher risk than former smokers (zero quit time, table 1). When the models are compared on equal grounds (same number of individuals to screen) the HUNT model outperforms both the NLST and the NELSON in all performance metrics employed (tables 4 and 5).

Moreover, in equal footing (to gain the same sensitivity as the NELSON or NLST) one would need screening 26.6% more people using the NELSON and 29.4% more people using the NLST criteria than the HUNT risk ranking (see Results section and figure 2). Sparing >25% of a heavy smoker population of screening designated as high-risk by NELSON and NLST criteria by using this

model could be considered cost effective. Given that the predictive performance of the HUNT model is superior in all metrics than the NLST and the NELSON criteria, implies that it is the most cost-effective of the three for a given risk threshold.

As we noted in our previous HUNT paper, and according to the original HUNT model, in a low-smoking population, one would need to screen only 22% of ever-smokers to identify 81.85% of all lung cancers within 6 years.[11] In this highly selected cohort, using the same threshold excludes very few from screening (3.2%) and identifies almost the whole DLCT cohort as high-risk, which is logical, since most were heavy smokers and current smokers. However, it predicts 149 out of 150 cases, 99.3%.

In contrast, applying the NLST selection criteria on the DLCST cohort is so strict that 53.8% of this population would be excluded and one would fail to screen 40.2% of the cases within this cohort (figure 1, table 3). Similarly, applying the NELSON study criteria on the DLCST, one would exclude from screening 17.4% of the population failing to identify 5.3% of the cancer cases (table 3).

A major difference between the HUNT model and fixed criteria sets, like the NELSON and the NLST, is that the former produces a ranking of individuals according to risk. To decide which individuals to screen, one needs to set a risk threshold above which screening should take place. The threshold should be set in the most cost-effective way, as a consensus, considering all public health factors (eg, screening capacity, risk due to screening, predictive performance of the model).

One may question whether these models are developed in a population with a representative spectrum of all lung cancer subtypes. The cancer subtypes in the DLCST were
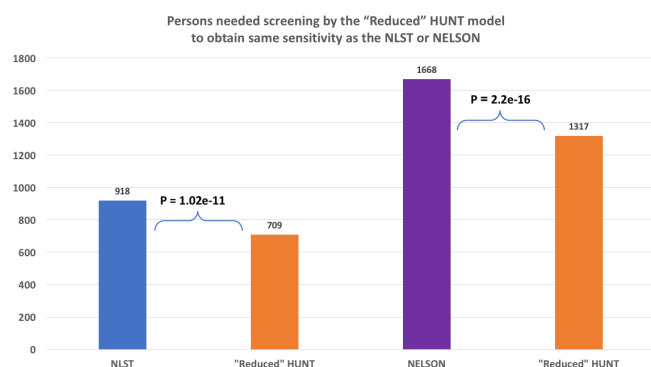

Figure 2 Comparison of persons needed screening by the 'Reduced' HUNT model to obtain same sensitivity as the NLST or NELSON in the unscreened population (n=1 999) (p values for comparing proportions were by χ² test, p<0.05 deemed significant).

Table 5 Direct comparison of the 'Reduced' HUNT model against NELSON criteria, for the same number of individuals (n=1668), on the control group of DLCST. The HUNT model shows increased predictive performance in all metrics (sensitivity, specificity, PPV, NPV)

	LC (N)	Without LC (N)	Total (N)	Predictive value
Population	52	1 947	1999	
'Reduced' HUNT*				
Criteria positive	50 TP (3.00%)	1618 FP (97.00%)	1668†	**PPV 3.00%**
Criteria negative	2 FN (0.60%)	329 TN (99.40%)	331	**NPV 99.40%**
Sensitivity	**96.15%**			
Specificity		**16.90%**		
NELSON‡				
Criteria positive	47 TP (2.82%)	1621 FP (97.18%)	1668*	PPV 2.82%
Criteria negative	5 FN (1.51%)	326 TN (98.45%)	331	NPV 98.45%
Sensitivity	90.38%			
Specificity		16.74%		

'Reduced' HUNT model.
*'Reduced' HUNT model criteria.
†The number of those picked by the NELSON, but with the top risk score by the 'Reduced' HUNT model.
‡NELSON criteria age 50–75, >15 cigarettes a day >25 years, or >10 cigarettes a day >30 years, both with 10 or less years quit time.
FN, false negative; FP, false positive; NPV, negative predictive value; PPV, positive predictive value; TN, true negative; TP, true positive.

previously published and include the whole spectrum of lung cancers, including 38% adenocarcinoma, 15% squamous cell lung cancer, 15% non-small cell lung cancer, 14% small-cell lung cancer and the rest broncho-alveolar carcinoma, large-cell carcinoma and combinations of histologies.[12] The original HUNT model was based on a population of >65 000 age 20–100 and since it was unselected, indeed cover the whole spectrum of lung cancer subtypes. Both HUNT models are therefore suitable for all types of lung cancers.

Comparison of HUNT with the PLCO$_{m2012}$ model

The five variables in this HUNT model were also found in another high-performing model, the PLCO$_{m2012}$ that consists of 10 variables, including COPD, family history of lung cancer and educational status.[16] The PLCO$_{m2012}$ was also found more sensitive than the NLST criteria in the NLST study, and showed how ranking of risk is superior to fixed criteria, like in our study. However, in contrast to the original HUNT model the PLCO$_{m2012}$ was developed in a population of age 55–74 and with a mean pack year count of ≈30 high, and thus light smokers with many years of smoking, people younger than 55 or older than 74 and as well as combinations of all those are not represented.[11] Moreover, there are two variables used in the PLCO$_{m2012}$ that are susceptible to bias, namely 'history of COPD' and 'family history of lung cancer', correspondingly reflected in biassed predictions. Specifically, in the case of COPD, misdiagnosis is very common : under- and over-diagnosis can be fivefold more common than correct diagnosis.[17] Similarly, 'family history of lung cancer' is also a variable that may be hard to obtain accurately as some people may not know details of their family history due to several reasons, such as estrangement, death, or adoption from

unknown donors.[16 18] Some studies showed that about one-quarter of those who had blood relatives with cancer documented in medical records did not report a family history. Such studies demonstrate the weakness and potential danger of transferring a population-based risk factor to a personal risk prediction model.[19 20]

Study limitations

The present study, analysis, and model exhibit several limitations. The original HUNT Lung Cancer Risk Model was developed in a total adult population of all degrees of smoking burden with 199 clinical variables to choose among, an ideal population to learn which are true, independent and interdependent risk factors.[11] Out of the 199 variables, 36 were manually selected based on expert knowledge; the data-driven feature selection ended up with seven clinical variables, out of which five were employed in the HUNT model presented here. The manual feature selection was performed so that the Cox model had enough statistical power to enable a backward feature selection methodology. However, one could potentially employ other feature selection method, more suitable to high-dimensional data, to perform feature selection directly with the original set of the 199 variables, possibly leading to better predictive models. In addition, other modelling techniques could be employed to try to improve the predictive power, for example, using Random Survival Forests, Support Vector Machines for censored time-to-event outcomes, and others.

To apply the HUNT model to a clinical setting, as mentioned above, one should choose a risk threshold for defining the high-risk population to screen. Such a threshold is not trivial to determine as it depends on several factors and public health issues.

A shortcoming of the present study is that the DLCST lacked two variables of the original HUNT model, namely, periodical daily cough and hours of exposure to indoor smoke. Thus, the current study serves only as a proxy of a direct comparison of the original HUNT model with the NLST and the NELSON criteria. Nevertheless, it is reasonable to expect that the inclusion of these extra sources of information would prove beneficial and not detrimental to the predictive power of the model.

Moreover, as much as we would like to compare the $PLCO_{m2012}$ model against the HUNT model on the Danish cohort, unfortunately, this is not possible as the $PLCO_{m2012}$ model requires variables not measured in that cohort. Specifically, the variables needed for the $PLCO_{m2012}$ model not included in the Danish trial are: race, education, COPD yes/no, personal history of cancer yes/no, family history of lung cancer yes/no. Nevertheless, an aggregate comparison between the original HUNT model and the PLCO, that is, a comparison on their average predictive metrics and not individual predictions on the same cohort, was performed in the HUNT original paper in EBiomedicine.[11]

Regarding the cohorts used in this study, both the training data (HUNT cohort) and the validation data (DLCST) are of European ancestry; it is thus untested, how well the model generalises to populations of other ancestries such as, African, Asian and South American that are under-represented in both cohorts. The generalisation of the model to populations that are either genetically or socially (lifestyle) quite different deserves further study. In addition, the training cohort lacked some quantities, not included in our original 36 considered variables. Clinical quantities such as significant comorbidities and employment exposures could possibly improve the predictive power of the model. Of course, molecular measurements could provide a wealth of predictive information that is not employed in the current work.

CONCLUSIONS

The HUNT model outperforms the NLST and the NELSON criteria in all predictive performance metrics, as demonstrated on the DLCST study, a screening trial in Denmark with a 10-year follow-up period. Unlike cut-off criteria regarding age, pack years and quit time, a risk prediction model ranks individuals according to a weighted average of many variables, permitting the medical community and public health authorities to determine the risk threshold that is most cost-effective for screening. This is the first study that directly compares favourably against the NELSON criteria. The results support the original HUNT Lung Cancer Model and indicate that this should be used prospectively in a screening study or programme. Data-driven, multivariable predictive models, when trained with state-of-the-art methods in statistics and machine learning, seem to outperform

current criteria and provide a potentially fruitful research direction that could be clinically important. Our hope is that public health authorities will consider this fact when determining screening guidelines.

Author affiliations

^{1}Department of Clinical and Molecular Medicine, Norges teknisk-naturvitenskapelige universitet, Trondheim, Norway
^{2}Cancer Clinic, Sykehuset Levanger, Levanger, Norway
^{3}Department of Computer Science, University of Crete - Voutes Campus, Heraklion, Greece
^{4}Institute of Applied Mathematics, Foundation for Research and Technology - Hellas (FORTH), Heraklion, Greece
^{5}Science and Technology Park of Crete, GNOSIS Data Analysis PC, Heraklion, Greece
^{6}Institute of Chemical Biology, Ilia State University, Tbilisi, Georgia
^{7}Department of Thoracic Surgery RT, Rigshospitalet, University of Copenhagen, Faculty of Health Sciences, Copenhagen, Denmark
^{8}Department of Respiratory Medicine, Gentofte University Hospital, Hellerup, Denmark
^{9}Department of Radiology, Akershus University Hospital, Lørenskog, Norway

ORCID iD
Oluf Dimitri Røe http://orcid.org/0000-0002-4870-5822

REFERENCES

1 Torre LA, Siegel RL, Ward EM, *et al*. Global cancer incidence and mortality rates and trends--an update. *Cancer Epidemiol Biomarkers Prev* 2016;25:16–27.
2 Aberle DR, Adams AM, Berg CD, *et al*. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011;365:395–409.
3 Pinsky PF, Berg CD. Applying the National lung screening trial eligibility criteria to the US population: what percent of the population and of incident lung cancers would be covered? *J Med Screen* 2012;19:154–6.
4 De Koning H, Van Der Aalst C, Ten Haaf K, *et al*. PL02.05 effects of volume CT lung cancer screening: mortality results of the Nelson

Randomised-Controlled population based trial. *J Thorac Oncol* 2018;13.

5. Walter JE, Heuvelmans MA, de Jong PA, *et al*. Occurrence and lung cancer probability of new solid nodules at incidence screening with low-dose CT: analysis of data from the randomised, controlled Nelson trial. *Lancet Oncol* 2016;17:907–16.

6. Oudkerk M, Devaraj A, Vliegenthart R, *et al*. European position statement on lung cancer screening. *Lancet Oncol* 2017;18:e754–66.

7. Ten Haaf K, Jeon J, Tammemägi MC, *et al*. Risk prediction models for selection of lung cancer screening candidates: a retrospective validation study. *PLoS Med* 2017;14:e1002277.

8. Katki HA, Kovalchik SA, Berg CD, *et al*. Development and validation of risk models to select Ever-Smokers for CT lung cancer screening. *JAMA* 2016;315:2300–11.

9. Wu X, Wen CP, Ye Y, *et al*. Personalized risk assessment in never, light, and heavy smokers in a prospective cohort in Taiwan. *Sci Rep* 2016;6:36482.

10. Muller DC, Johansson M, Brennan P. Lung cancer risk prediction model incorporating lung function: development and validation in the UK Biobank prospective cohort study. *J Clin Oncol* 2017;35:861–9.

11. Markaki M, Tsamardinos I, Langhammer A, *et al*. A validated clinical risk prediction model for lung cancer in smokers of all ages and exposure types: a HUNT study. *EBioMedicine* 2018;31:36–46.

12. Wille MMW, Dirksen A, Ashraf H, *et al*. Results of the randomized Danish lung cancer screening trial with focus on high-risk profiling. *Am J Respir Crit Care Med* 2016;193:542–51.

13. Krokstad S, Langhammer A, Hveem K, *et al*. Cohort profile: the HUNT study, Norway. *Int J Epidemiol* 2013;42:968–77.

14. Harrell FE. *Regression modeling strategies : with applications to linear models, logistic regression, and survival analysis*. New York: Springer, 2001.

15. Royston P, Altman DG. External validation of a COX prognostic model: principles and methods. *BMC Med Res Methodol* 2013;13:33.

16. Weber M, Yap S, Goldsbury D, *et al*. Identifying high risk individuals for targeted lung cancer screening: Independent validation of the PLCO$_{m2012}$ risk prediction tool. *Int J Cancer* 2017;141:242–53.

17. Gershon AS, Thiruchelvam D, Chapman KR, *et al*. Health services burden of undiagnosed and overdiagnosed COPD. *Chest* 2018;153:1336–46.

18. Marcus MW, Chen Y, Raji OY, *et al*. LLPi: Liverpool lung project risk prediction model for lung cancer incidence. *Cancer Prev Res* 2015;8:570–5.

19. Glanz K, Grove J, Le Marchand L, *et al*. Underreporting of family history of colon cancer: correlates and implications. *Cancer Epidemiol Biomarkers Prev* 1999;8:635–9.

20. Ozanne EM, O'Connell A, Bouzan C, *et al*. Bias in the reporting of family history: implications for clinical care. *J Genet Couns* 2012;21:547–56.