Predicting the future risk of lung cancer

Development and validation of QCancer2 (10-year risk) lung model and evaluating the performance of nine prediction models

Weiqi Liao (presenting), Carol Coupland, Judith Burchardt, David Baldwin, the DART initiative (WP6), Fergus Gleeson, Julia Hippisley-Cox

Research background and rationale

❖ Lung cancer – the most common cause of cancer death in the UK (21%)

❖ Screening helps early detection and reduce mortality.

❖ Using risk prediction models can target individuals at high risk.

❖ The Liverpool Lung Project (LLP$_{v2}$) model (UK) and the Prostate, Lung, Colorectal and Ovarian (PLCO$_{m2012}$) model (US), currently used in the Targeted Lung Health Check Programme, did not have satisfactory model performance.

❖ This study aims to develop and validate a model to predict the future risk of lung cancer and suitable for lung cancer screening in the English primary care population.

Conclusions

❖ Compared with the currently used LLP$_{v2}$ and PLCO$_{M2012}$ models in the Targeted Lung Health Check Programme, the QCancer2 (10-year risk) lung model has better
  • Discrimination
  • Calibration
  • Net benefit

❖ The QCancer2 (10-year risk) lung model may be more suitable for selecting individuals at high risk from the English primary care population for lung cancer screening.

Table 1 – Basic information of the nine models

Prediction models	QCancer2 (10-year) lung model	LLP	LCRAT	PLCO	Pittsburgh	Bach
Model versions	2015, 2022	v2, v3		2012, 2014		
Country / population	England	England	US	US	US	US
Predictive horizon	1-10 years	5 years	5 years	6 years	6 years	10 years
Age range (years)	25-84	40-84	55-74	55-74	50-79	45-69
Includes never-smokers	Yes	Yes	No	Only the 2014 model	No	No

Figure 1 – Examples of calibration plots (validation, 910,870 women ever-smoker aged 55-74 years old)

5-year predictive horizon – LLP$_{v2}$ VS QCancer2
10,720 lung cancer cases

6-year predictive horizon – PLCO$_{M2012}$ VS QCancer2
12,595 lung cancer cases

Box 1
Predictors for the QCancer 2 (10-year risk) lung model

Sociodemographic:
• Age
• Sex
• Ethnicity
• Socioeconomic status (Townsend score)

Lifestyle factors:
• Smoking status
• Smoking intensity (cigarette per day)
• Alcohol
• BMI

Comorbidities:
• COPD
• Asthma
• History of pneumonia
• Venous thromboembolism
• Asbestos exposure
• Personal history of cancer
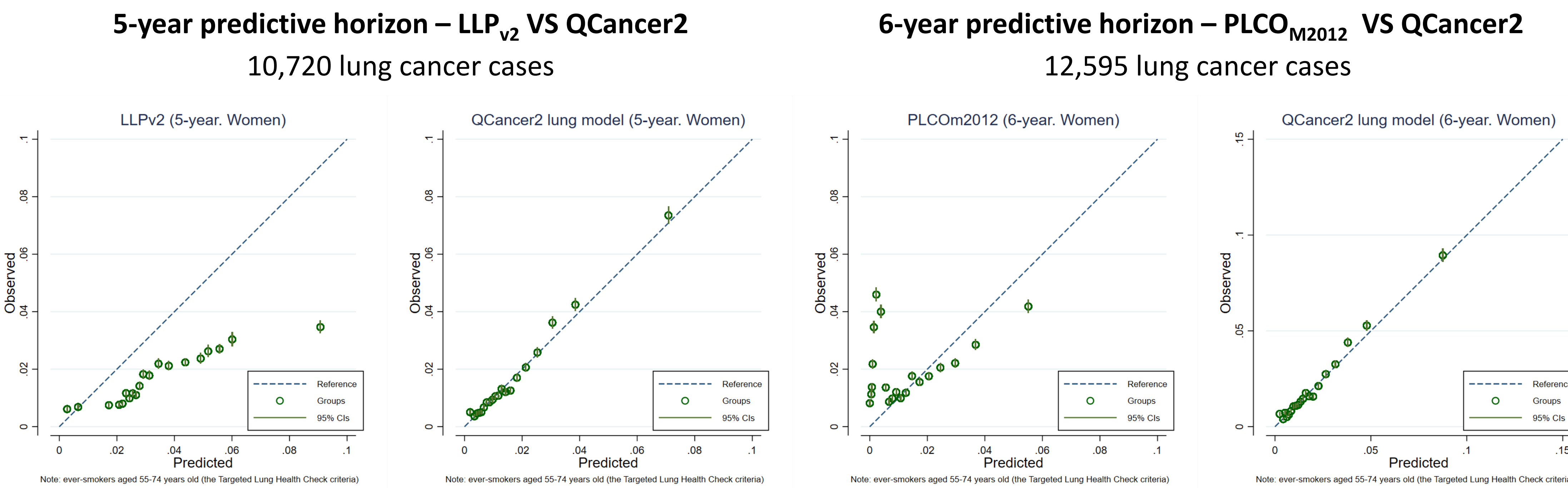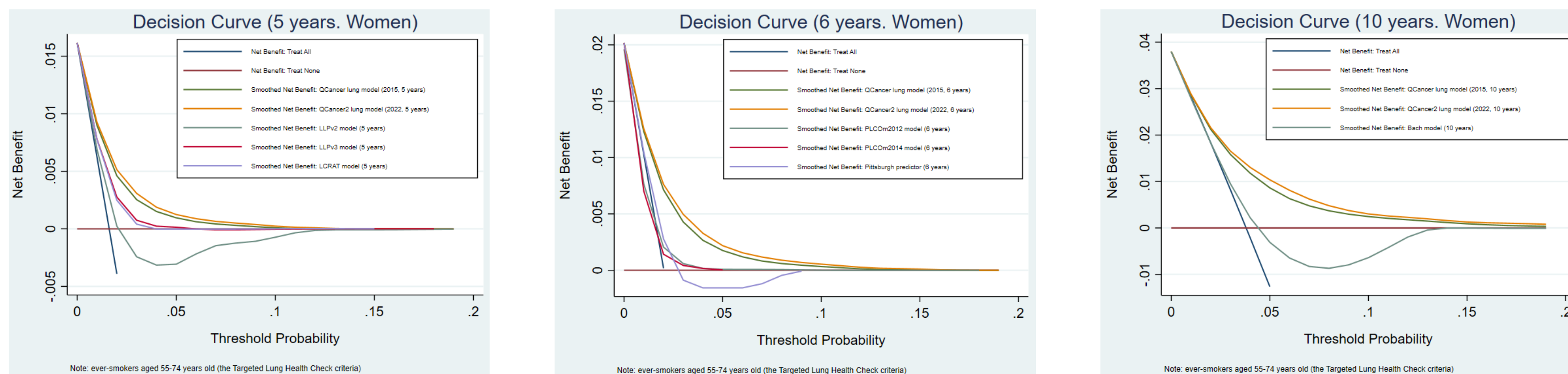• Family history of lung cancer

Figure 2 – Decision curve analysis (910,870 women ever-smoker aged 55-74 years old)

Note: similar results between men and women. This poster only presents the results of women due to limited space.

Methods

Stage 1 – Develop and validate the QCancer2 (10-year risk) lung model

❖ *Model development*: Cox regression was used in the derivation dataset (12.99 million) to develop the QCancer2 (10-year risk) lung model in men and women separately, using data from the QResearch® database.
  • Multiple imputation was used to replace missing values (5 imputations).
  • Fractional polynomials [2] were used to model non-linear relationships between age/BMI/Townsend scores and the outcome (incident diagnosis of lung cancer).

❖ *Model validation*: three discrimination measures (Harrell's C [3], D statistic [4], R_D^2 [5]) and calibration plots were used to evaluate the model performance in the validation cohort (4.14 million).

Stage 2 – Model evaluation

❖ The QCancer2 (10-year risk) lung model was compared with the other seven models (LLP$_{v2}$, LLP$_{v3}$, LCRAT, PLCO$_{M2012}$, PLCO$_{M2014}$, Pittsburgh, and Bach models) to predict incident lung cancer diagnosis in two approaches:
  1) In current and ex-smokers aged 55-74 years (the population of the Targeted Lung Health Check Programme),
  2) The QCancer2 lung model compared with each model using its eligibility criteria for the study sample/population.

❖ Model performance was evaluated by discrimination and calibration plots.

❖ Decision curve analysis [6] was used to evaluate the net benefit.

References:
1. O'Dowd et al. Thorax. 2022;77(9):882-890;   2. Royston et al. Int J Epidemiol 1999;28(5):964-74.
3. Newson. Stata Journal 2010;10(3):339-58;   4. Royston el al. Stat Med 2004;23:723-48;
5. Royston. Stata J 2006;6:1-14;   6. Vickers et al. Med Decis Making 2006;26(6):565-74.

Table 2 – Discrimination statistics of prediction models in women in the full model and ever-smokers aged 55-74 years old

	Harrell's C	D statistic	R_D^2
QCancer2 full model			
25-84 years old	0.90	2.81	65.4%
Predictive horizon: 5 year	Ever-smokers	55-74 years	
QCancer2	0.73	1.93	46.9%
LLP$_{v2}$	0.65	1.56	36.7%
LLP$_{v3}$	0.66	1.63	38.9%
LCRAT	0.64	1.57	37.1%
Predictive horizon: 6 year			
QCancer2	0.73	1.91	46.5%
PLCO$_{M2012}$	0.53	1.28	28.1%
PLCO$_{M2014}$	0.53	0.73	11.4%
Pittsburgh	0.64	1.58	37.3%
Predictive horizon: 10 year			
QCancer2	0.72	1.90	46.2%
Bach	0.58	1.35	30.2%

NUFFIELD DEPARTMENT OF
PRIMARY CARE
HEALTH SCIENCES

QRESEARCH   DART   UKRI   Innovate UK   NIHR | National Institute for Health and Care Research

@WLiao_Ox @JuliaHCox @dartlunghealth   Email: Weiqi.Liao@phc.ox.ac.uk