Biomarkers in lung cancer screening: the importance of study design

David R. Baldwin[1,2], Matthew E. Callister[3,4], Philip A. Crosbie[5,6], Emma L. O'Dowd[1,2], Robert C. Rintoul[7,8], Hilary A. Robbins[9] and Robert J.C. Steele[10]

Affiliations: [1]Respiratory Medicine, Nottingham University Hospitals, Nottingham, UK. [2]University of Nottingham, Nottingham, UK. [3]Leeds Teaching Hospitals, Leeds, UK. [4]University of Leeds, St James's University Hospital, Leeds, UK. [5]Division of Infection, Immunity and Respiratory Medicine, University of Manchester, Manchester, UK. [6]Manchester Thoracic Oncology Centre, North West Lung Centre, Manchester University NHS Foundation Trust, Manchester, UK. [7]Dept of Oncology, University of Cambridge, Cambridge, UK. [8]Dept of Thoracic Oncology, Royal Papworth Hospital, Cambridge, UK. [9]Genetic Epidemiology Group, International Agency for Research on Cancer, Lyon, France. [10]UK National Screening Committee, Dept of Surgery, Ninewells Hospital, University of Dundee, Dundee, UK.

Correspondence: David R. Baldwin, University of Nottingham, Epidemiology and Public Health, City Hospital Campus, Nottingham NG5 1PB, UK. E-mail: david.baldwin@nottingham.ac.uk

@ERSpublications
Use of biomarkers to select participants for lung cancer CT screening could improve clinical and cost effectiveness. It is crucial to design research that evaluates biomarkers against LDCT and to test them in the population in which they will be used. https://bit.ly/3gPHAPu

Cite this article as: Baldwin DR, Callister ME, Crosbie PA, *et al*. Biomarkers in lung cancer screening: the importance of study design. *Eur Respir J* 2021; 57: 2004367 [https://doi.org/10.1183/13993003.04367-2020].

In most randomised controlled trials of low radiation dose computed tomography (LDCT) screening for lung cancer, eligibility has been determined by age and tobacco smoking criteria [1, 2]. A number of multivariable risk prediction models have been developed that are more sensitive and specific, but are still heavily dependent on smoking and age [3]. Some have been used successfully in trials and pilot programmes and have yielded higher detection rates, although they may also select people with more comorbidities [4–6]. Using biomarkers to select people for lung cancer screening is an important area for research because it offers the possibility to identify those at high risk who are not eligible for screening by conventional criteria, while avoiding further investigation in those who do not have malignant disease. This means that the impact of the intervention as a whole could be greater (more lives saved) and harms, due to the screening of low-risk people or those without cancer, could be avoided. Two biomarkers intended for use in lung screening are presented in this issue of the *European Respiratory Journal*.

Sullivan *et al.* [7] describe the use of the "EarlyCDT-Lung", a seven-autoantibody blood test to detect lung cancer which was evaluated in a randomised trial in Scotland (Early Lung Cancer Detection in Scotland, ECLS). Just over 12 000 people who were at high risk of developing lung cancer by age and smoking criteria were randomised 1:1 to an intervention arm or usual care. In the intervention arm, participants were offered EarlyCDT-Lung, and those who were test-positive (~10%; n=598) went on to have an initial chest radiograph followed by LDCT every 6 months for 24 months. The researchers report a lower proportion of late stage cancers in the active arm, 33 of 56 (59%) compared with 52 of 71 (73%) in the control arm (hazard ratio 0.64, 95% CI 0.41–0.99). Seven patients with unclassified stage in the control arm were assumed to be diagnosed at late stage, whereas all in the intervention arm had known stage. This assumption favoured the primary outcome of the study, making the difference in late stage disease

Received: 30 Nov 2020 | Accepted after revision: 7 Dec 2020

statistically significant. The authors reported sensitivity estimates 2 years after recruitment, for lung cancer overall and stratified by early/late stage disease.

The most important limitation in the research design, which the researchers acknowledge, was that neither the participants who tested negative by EarlyCDT-Lung, nor the participants in the control arm, were offered LDCT. This has three implications for interpreting the results presented. Firstly, the ability of LDCT screening to diagnose lung cancer at earlier stages is well-established, with 75% or more diagnosed at stage I/II [1, 2, 5]. Since LDCT screening was only offered to those who tested positive by the biomarker, it is not possible to determine whether EarlyCDT-Lung made any improvement to earlier detection. Indeed, a test generating a result at random would be expected to result in earlier diagnosis in the intervention group, simply by triggering a CT scan in a proportion of these participants.

Secondly, assessing overall sensitivity for lung cancer after only 2 years of follow-up is likely to overestimate the sensitivity of the EarlyCDT-Lung screening intervention (reported as 32%). This is because lung cancer has a sojourn time (interval between development and presentation of cancer) of 4–6 years [8]; indeed, "catch-up" of cancer presentations in the non-LDCT-screened group in the National Lung Screening Trial continued for 9 years [9]. Therefore, one would expect continued identification of lung cancers in the test-negative group over subsequent years of follow-up, thereby progressively reducing the sensitivity (*i.e.* the proportion identified by the test). A more accurate estimate of the sensitivity of the EarlyCDT-Lung would have been obtained if all participants in the screened arm had received an exit LDCT, because LDCT detects many of these cancers. Furthermore, a strange aspect of the ECLS results is that more lung cancers were recorded in the control group (n=71) than in the intervention group (n=56). LDCT typically detects a 2.2-fold (NELSON), 2.4 (MILD) to 3.1-fold (LUSI) greater number of cancers in the LDCT arm compared with the no-screening control arm in the first 2 years [2, 10, 11]. In the ECLS, 71 lung cancers were detected over 2 years in the control arm. If LDCT had been offered to all intervention-arm participants we would expect at least 156 (2.2×71) lung cancers to be detected. Among these, approximately 117 (75%) would be expected to be stage I/II. Since only 12 participants had stage I/II lung cancer in the EarlyCDT-Lung positive group, the test sensitivity would be 10%. The ECLS reported sensitivity of 52% for stage I/II disease is, therefore, likely to be a substantial over-estimate and entirely due to the fact that only test-positive partcipants had LDCT.

Finally, measuring sensitivity for early *versus* late stage disease (in the absence of screening for the test-negative group) is flawed. Any comparison of sensitivities between these two groups is biased by the fact that the cancers in the test positive group will generally be early stage cancers detected by LDCT screening, whereas those in the test-negative participants will present at a later time-point and are therefore likely to be more advanced stage. Thus, EarlyCDT-Lung must be compared directly with existing methods for selecting people for screening (and should include LDCT for test negative participants) before any conclusions about its utility can be drawn. At worst, it could cause considerable harm by excluding people at high risk who would be selected by currently available and validated multivariable models. It is noted that in ECLS, 90% of participants were effectively excluded from LDCT by virtue of the negative blood test.

A number of other biomarkers for lung screening are in various stages of development [12]. However, biomarkers with promising early results have often failed in subsequent research [13, 14], and this is often explained by differences in design of studies between development and validation. For initial development studies, biomarkers are typically tested for differences between cases and controls. A common pitfall is that biomarkers are capable of distinguishing between cases and controls at the time of diagnosis, but fail when tested in pre-diagnostic samples, or fail in the setting of early stage disease. Another pitfall is that cases and controls are often poorly matched for important factors, or even derived from separate populations. When later validated in a population of at-risk people, the test performs less well. Therefore, initial control populations should be derived from the same cohort as the patients with cancer and selected in a way that simulates the intended downstream use of the biomarker. For example, biomarkers intended for use in selection for screening should be tested in a group potentially eligible for screening. The biomarker should then be compared with the gold-standard method for the clinical application of interest, for example, to risk prediction models if intended for use in defining eligibility for screening.

Also in this issue of the *ERJ*, Gaga *et al.* [15] report on a novel methylation-based biomarker, Lung EpiCheck, which shows a sensitivity of 85% for early stage lung cancer, with a specificity of 64% when combined with other risk factors. This new test is promising but, quite rightly, the authors recommend further prospective trials to establish how the test performs in populations eligible for screening. As the authors acknowledge, there were important differences in the cases and controls in the European dataset due to retrospective selection of cases and identification of controls from an outpatient population. Furthermore, the controls did not have follow-up or LDCT to look for early stage lung cancer. In the

single centre Chinese validation set, healthy controls were used. Thus, the high AUC of 0.94 is likely a result of the marked difference between cases and controls and will reduce when the biomarker is used to select people for screening where cases and controls are from the same population. One important step will be to use the test alongside LDCT screening and compare the performance with that of currently validated risk models.

Biomarkers offer a promising approach to improve the impact of screening, because unlike conventional risk prediction models, they may be able to identify people who have high risk for lung cancer without concurrently elevated risk for competing causes of death related to age and smoking. However, it is important to establish whether a positive biomarker is also a marker for poor outcome, as biomarkers may detect more aggressive disease with a poor prognosis [16]. For example, some cell-free DNA tests may be correlated with more aggressive tumours [17–19]. Conventional cancer screening reduces mortality mainly by detecting less aggressive tumours (length time effect) so although detecting aggressive cancers early may improve treatment options, the impact on mortality may be less. Sullivan *et al.* [7] show that a positive EarlyCDT-Lung, based on measuring the immune response to lung cancer, in combination with LDCT showed a non-significant trend to reduced mortality. This may merit further study.

The ultimate aim is to have a biomarker that is inexpensive and can be used as the screening test, with LDCT being the downstream investigation, in a similar way to colorectal screening where the faecal immunochemical test or faecal occult blood test is followed by colonoscopy. It is important that the test identifies not only people who will develop disease, but specifically people who can benefit.

Many promising biomarkers are emerging [20–22], but the onus is on the research community to test each one using robust methodology. Studies must be designed in such a way that the added contribution of the biomarker can be clearly identified and quantified, in direct comparison to the gold standard method for the task that the biomarker purports to accomplish. This is an exciting and promising field, but the excitement must be met with carefully designed, definitive studies.

Where authors are identified as personnel of the International Agency for Research on Cancer/World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer/World Health Organization.

Conflict of interest: D.R. Baldwin reports personal fees from AstraZeneca, MSD, BMS, Roche and Johnson and Johnson, outside the submitted work. M.E. Callister has nothing to disclose. P.A. Crosbie reports personal fees from AstraZeneca and Novartis, personal fees and other from Everest Detection, outside the submitted work. E.L. O'Dowd has nothing to disclose. R.C. Rintoul reports personal fees from AstraZeneca and Roche, outside the submitted work; and is Chief Investigator of the LuCiD study, sponsored by Owlstone Medical but declares no financial support or financial interest in Owlstone Medical. H.A. Robbins has nothing to disclose. R.J.C. Steele has nothing to disclose.

Support statement: H.A. Robbins was supported in part by the INTEGRAL programme (U19 CA203654). P.A. Crosbie was supported by the NIHR Manchester Biomedical Research Centre. R.C. Rintoul is supported by the NIHR Cambridge Biomedical Research Centre and Cancer Research UK Cambridge Centre.

References

1 National Lung Screening Trial Research Team, Aberle DR, Adams AM, *et al.* Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011; 365: 395–409.
2 de Koning HJ, van der Aalst CM, de Jong PA, *et al.* Reduced lung-cancer mortality with volume CT screening in a randomized trial. *N Engl J Med* 2020; 382: 503–513.
3 Ten Haaf K, Jeon J, Tammemagi MC, *et al.* Risk prediction models for selection of lung cancer screening candidates: a retrospective validation study. *PLoS Med* 2017; 14: e1002277.
4 Tammemagi MC, Schmidt H, Martel S, *et al.* Participant selection for lung cancer screening by risk modelling (the Pan-Canadian Early Detection of Lung Cancer (PanCan) study): a single-arm, prospective study. *Lancet Oncol* 2017; 18: 1523–1531.
5 Field JK, Duffy SW, Baldwin DR, *et al.* UK Lung Cancer RCT Pilot Screening Trial: baseline findings from the screening arm provide evidence for the potential implementation of lung cancer screening. *Thorax* 2016; 71: 161–170.
6 Mazzone PJ, Silvestri GA, Patel S, *et al.* Screening for lung cancer: CHEST guideline and expert panel report. *Chest* 2018; 153: 954–985.
7 Sullivan FM, Mair FS, Anderson W, *et al.* Earlier diagnosis of lung cancer in a randomised trial of an autoantibody blood test followed by imaging. *Eur Respir J* 2021; 57: 2000670.
8 Ten Haaf K, van Rosmalen J, de Koning HJ. Lung cancer detectability by test, histology, stage, and gender: estimates from the NLST and the PLCO trials. *Cancer Epidemiol Biomarkers Prev* 2015; 24: 154–161.
9 National Lung Screening Trial Research Team. Lung cancer incidence and mortality with extended follow-up in the National Lung Screening Trial. *J Thorac Oncol* 2019; 14: 1732–1742.
10 Paci E, Puliti D, Lopes Pegna A, *et al.* Mortality, survival and incidence rates in the ITALUNG randomised lung cancer screening trial. *Thorax* 2017; 72: 825–831.
11 Becker N, Motsch E, Trotter A, *et al.* Lung cancer mortality reduction by LDCT screening: results from the randomized German LUSI trial. *Int J Cancer* 2020; 146: 1503–1513.

12 Seijo LM, Peled N, Ajona D, *et al.* Biomarkers in lung cancer screening: achievements, promises, and challenges. *J Thorac Oncol* 2019; 14: 343–357.
13 Ilie M, Hofman V, Long-Mira E, *et al.* "Sentinel" circulating tumor cells allow early diagnosis of lung cancer in patients with chronic obstructive pulmonary disease. *PLoS One* 2014; 9: e111597.
14 Marquette CH, Boutros J, Benzaquen J, *et al.* Circulating tumour cells as a potential biomarker for lung cancer screening: a prospective cohort study. *Lancet Respir Med* 2020; 8: 709–716.
15 Gaga M, Chorostowska-Wynimko J, Horváth I, *et al.* Validation of Lung EpiCheck, a novel methylation-based blood assay, for the detection of lung cancer in European and Chinese high-risk individuals. *Eur Respir J* 2021; 57: 2002682.
16 Yuxia M, Zhennan T, Wei Z. Circulating miR-125b is a novel biomarker for screening non-small-cell lung cancer and predicts poor prognosis. *J Cancer Res Clin Oncol* 2012; 138: 2045–2050.
17 Gressel GM, Maggi EC, Harmon BE, *et al.* Low molecular weight serum cell-free DNA concentration is associated with clinicopathologic indices of poor prognosis in women with uterine cancer. *J Transl Med* 2020; 18: 323.
18 Lennon AM, Buchanan AH, Kinde I, *et al.* Feasibility of blood testing combined with PET-CT to screen for cancer and guide intervention. *Science* 2020; 369: eabb9601.
19 Cohen JD, Li L, Wang Y, *et al.* Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science* 2018; 359: 926–930.
20 Shen SY, Singhania R, Fehringer G, *et al.* Sensitive tumour detection and classification using plasma cell-free DNA methylomes. *Nature* 2018; 563: 579–583.
21 Liu MC, Oxnard GR, Klein EA, *et al.* Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. *Annals of Oncology* 2020; 31: 745–759.
22 Chabon JJ, Hamilton EG, Kurtz DM, *et al.* Integrating genomic features for non-invasive early lung cancer detection. *Nature* 2020; 580: 245–251.