EXPERT REVIEWS

Use of lung cancer risk models in planning research and service programs in CT screening for lung cancer

Expert Rev. Anticancer Ther. 9(10), 1467–1472 (2009)

Stephen W Duffy†, Olaide Y Raji, Olorunsola F Agbaje, Prue C Allgood, Adrian Cassidy and John K Field

†*Author for correspondence*
Cancer Research UK Centre for EMS, Wolfson Institute of Preventive Medicine, Barts and the London School of Medicine and Dentistry, Charterhouse Square, London EC1M 6BQ
Tel.: +44 207 882 3535
Fax: +44 207 882 0269
s.w.duffy@qmul.ac.uk

Computed tomography screening for lung cancer is now being tested in a number of international trials. The long-term success of the approach in the future National Screening Programme is dependent upon identifying populations at sufficient risk of lung cancer that the benefit–harm ratio of the intervention is likely to be high. There are a number of lung cancer risk prediction models currently available. We review these, and demonstrate, using the Liverpool Lung Project risk prediction model as a case study, the potential for use of a risk prediction model in the design of a randomized trial of lung cancer screening and in the planning of a service screening program.

Keywords: computed tomography • lung cancer • risk model • screening • trial design

In recent years, the epidemiological research effort has focused on predicting an individual's absolute risk of various chronic diseases, to identify high-risk groups that may benefit from surveillance or preventive interventions [1–3]. The advent of computed tomography (CT) screening for lung cancer, which is currently being evaluated in large-scale randomized studies, implies a need to identify suitably high-risk populations that may benefit from the screening [4]. Consequently, the development of a predictive model for an individual's absolute risk of lung cancer has become a topical issue [5–8].

Unusually in chronic disease epidemiology, lung cancer is characterized by a major risk factor, cigarette smoking, which in itself accounts for the majority of cases of the disease [9]. However, the predictive ability of other identified risk factors, notably familial or genetic factors, is highly topical for a number of reasons [7,8,10,11]:

- As smoking rates decline, particularly in the developed world, nontobacco risk factors of lung cancer gain in relative importance;
- Even in a population of never-smokers, lung cancer would still represent a considerable public-health burden;
- A majority of long-term smokers do not develop lung cancer; identifying the reasons for this could have major implications for prevention and control.

Lung cancer risk models

Several individual risk models for lung cancer have been developed. Van Klaveren and colleagues described a model based only on age and smoking history [5], partly derived from the findings of Peto *et al.*, which they use to identify risk groups for inclusion in a randomized control trial (RCT) of CT screening for lung cancer [9]. The model stratifies the population into discrete groups with respect to age and duration of smoking, forming a two-way table with a risk score for each cell in the table. This has the advantage of simplicity of use and interpretation, although it will inevitably fail to identify some subjects who are at high risk due to risk factors that do not pertain to smoking [7,8].

Bach and colleagues developed models of lung cancer incidence and mortality, based on age, sex, smoking history and asbestos exposure, specifically for use in populations already known to be at elevated risk of lung cancer [6]. The model was developed from prospective cohort data, and has been used to estimate the

 10.1586/ERA.09.87 ISSN 1473-7140

likely effect of CT screening in a number of uncontrolled studies, by comparing observed deaths from lung cancer with the numbers expected under the model [12]. Although some tactical aspects of the analysis and interpretation of the latter exercise have been criticized [13], the overall strategy of risk prediction and observed/expected comparison for single-arm studies is a powerful and attractive one.

Spitz and colleagues presented a set of three models of risk of lung cancer, for never-smokers, current smokers and ex-smokers, respectively, based on a case–control study with 1851 lung cancer cases and 2001 controls [7]. For never-smokers, the model includes exposure to environmental tobacco smoke and family history of lung cancer. For ex-smokers, the model includes exposure to a variety of dusts, history of hay fever (associated with a reduced risk), history of emphysema, family history of any cancer and age at smoking cessation. For current smokers, the model includes history of emphysema, pack-years of smoking, family history of smoking-related cancers, asbestos exposure and history of hay fever (associated with a reduced risk). The models gave 1-year predicted probabilities of lung cancer based on these factors. Validation exercises gave areas under the ROC curve of approximately 60%.

The Liverpool Lung Project (LLP) risk model was based on a case–control study of 597 cases and 1179 controls [8]. As well as accounting for the three most important risk factors of lung cancer – age, sex and smoking – the model incorporates other important factors, such as history of pneumonia, history of nonlung malignancy, asbestos exposure and family history of lung cancer, including age at onset of disease in affected relatives. Using the following mathematical expression, the LLP risk model provides an estimated absolute risk of lung cancer within the next 5 years for an individual with a specified combination of risk factors.

$$P = \frac{1}{1 + \exp(-\alpha - \Sigma\beta_i z_i)}$$

Where P is the probability of lung cancer in the next 5 years, z_i are the risk factors (e.g., smoking duration) and β are the logistic regression coefficients calculated from the case–control data. The α term is calculated using locally derived age and sex-specific incidence rates of lung cancer. Details of the calculations are provided in Cassidy *et al.* [8]. The βs are shown in Table 1.

As an illustration, consider the example of a male aged 67 years, with no smoking or pneumonia history, but with a family history of lung cancer in a relative aged under 60 years at diagnosis, personal history of another malignancy and a history of working with asbestos. For a man aged 67 years, the α is calculated as -5.435 [8]. The probability of lung cancer in the next 5 years is estimated as:

$$P = \frac{1}{1 + \exp(5.435 - 0.703 - 0.675 - 0.634)} = 0.0316$$

That is, this man is estimated to have a 3.16% risk of lung cancer in the next 5 years.

Now consider a man aged 64 years with 42 years of smoking, a family history of lung cancer with the affected relative aged 60 years or over at diagnosis, a personal history of another malignancy, but no asbestos exposure or history of pneumonia. The α is calculated as -5.601 [8]. His 5-year probability of lung cancer is:

$$P = \frac{1}{1 + \exp(5.601 - 2.507 - 0.675 - 0.168)} = 0.0953$$

That is, his risk of lung cancer in the next 5 years is 9.53%.

The internal tenfold cross-validation of the LLP risk model showed an average area under the ROC of 70%. Both the Spitz and the LLP models are currently undergoing external validation.

Table 1. Components of the Liverpool Lung Project multivariate risk model, calculated from the Liverpool Lung Project case–control study.

Risk factor	Category	Odds ratio	95% CI	Model coefficient
Duration of smoking (years)	Never smoked	1.00		0
	1–20	2.16	1.21–3.85	0.769
	21–40	4.27	2.62–4.94	1.452
	41–60	12.27	7.41–20.30	2.507
	>60	15.25	5.71–40.65	2.724
History of pneumonia	No	1.00		0
	Yes	1.83	1.26–2.64	0.602
Asbestos exposure	No	1.00		0
	Yes	1.89	1.35–2.62	0.634
Previous other malignancy	No	1.00		0
	Yes	1.96	1.22–3.14	0.675
Family history of lung cancer	No	1.00		0
	Yes, <60 years	2.02	1.18–3.45	0.703
	Yes, 60+ years	1.18	0.79–1.76	0.168

Use of risk models in design of a CT screening trial

Computed tomography screening for lung cancer is a hotly debated topic [12,13]. A number of RCTs of the CT screening for lung cancer are underway [4], and there is an ongoing feasibility study in the UK for a trial of CT screening versus no screening [101]. It is important to define an appropriate high-risk group for such a trial because the intervention will have harms as well as potential benefits [14], and the study population should be at sufficient risk to give a high potential benefit–harm ratio. Secondly, the risk should be high enough

that a trial of feasible size and duration will have sufficient statistical power to identify the effect of the intervention as significant.

Mortality is the primary end point in a RCT of CT screening for lung cancer. Therefore, the trial needs to be powered for a comparison of lung cancer mortality in the study and control groups. The number of lung cancer deaths in the study and control groups will depend on the study population incidence, the mean sojourn time (the duration of the phase where the tumor is asymptomatic but detectable by screening), the survival of screen-detected and symptomatic lung cancers, the compliance with screening in the study group and the sensitivity of the screening test. The mathematical expressions for the number of lung cancer death are available from the first author. We take our estimates of mean sojourn time of 2.06 years and sensitivity of 96% from Chien and Chen's review, which draws together the available evidence on screen-detection and interval cancer rates from six studies, and uses established Markov process methodology for estimation of mean sojourn time [15]. The 5-year survival of symptomatic disease in the UK is estimated at 6% [16] and, to be conservative, we assume 5-year survival of screen-detected cancers to be 50%, considerably lower than has been reported [17]. We assume 80% compliance with the invitation to screening in the study group.

If we stipulate a minimum risk of lung cancer to be 1.5% over 5 years, the average risk in those qualifying would be approximately 2.5%, as estimated from the LLP. There is particular interest in whether a single- or a repeat-screen study would be more economical. Assuming a risk of 2.5%, the relative risk of lung cancer death (study vs control) with a single screen in the study group and only follow-up thereafter, and follow-up alone in the control group, and numbers of subjects needed per group for 90% power, by year since randomization is shown in Table 2. This shows that the optimum time of analysis would be 3 years after randomization, and would require 25,000 subjects per group. The actual time after the start of the study would be more likely to be 5 years, owing to the fact that recruitment and randomization of 50,000 subjects would take several years.

The corresponding figures for a study with three annual screens and follow-up thereafter (Table 2) reveal that the optimum time of analysis would be 5 years after randomization with a size of 10,000 subjects per group required for the same 90% power. This might be amended to 6 years to allow for a shorter recruitment period required for 20,000 subjects. On this basis, one might consider the three-screen design to be more efficient. However, it would require 3 × 10,000 = 30,000 screening appointments and 24,000 actual CT screens, given the 80% compliance rate, whereas the single-screen design would require only 25,000 appointments and 20,000 screens. In addition, the single-screen design is positioned to answer the research question approximately 1 year earlier.

It is also possible that even with our addition of time to take account of the recruitment period, a longer follow-up will be required than anticipated, possibly due to failures of assumptions regarding sojourn time or survival. However, the same delays would be likely to apply to both design strategies. It should also be noted that in the event of the randomized trials showing a benefit of screening, the service screening programs that might follow would have longer durations of screening than 1 or 3 years. For example, breast cancer screening programs are typically offered for two decades of life.

We might consider a 1.5% minimum risk of lung cancer to be too low. If we chose a 2.5% minimum 5-year risk, the LLP control group indicate that the average risk in the study population would be between 3.5 and 4%. Again, we assume 3.5% to be conservative. Table 3 gives the corresponding relative risks and numbers required for 90% power. The single-screen design would give an answer at 3 years (realistically 5 years) with 16,000 subjects per arm and the three-screen design would report at 5 years (realistically 6 years) with 7000 per arm. Again, the single-screen design might be regarded as more cost effective, requiring only 12,800 CT scans compared with 16,800 required in the three-screen design.

Note that the underlying risk is a crucial ingredient of the design calculations. Changing the minimum risk criterion from 1.5% over 5 years to 2.5% reduces the required sample size by a factor of approximately a third.

We assumed a conservative estimate of survival of screen-detected cases (50%), to allow for possible length bias or overdiagnosis. If we had used the quoted survival, which is in excess of 80% [17], our study size would have been more than halved. In view of the evidence of overdiagnosis, even from chest x-ray [18], it would probably be over optimistic to do so.

Use of the risk model in designing a population screening program

If the lung cancer mortality benefit is confirmed in ongoing RCTs of CT screening for lung cancer, healthcare providers would be compelled to consider offering a screening program as a service, including that of the target population. As noted

Table 2. Relative risks of lung cancer death and required number per group by year for a 5-year risk of 2.5%.

Year	One annual screen		Three annual screens	
	RR (intervention vs control)	*Required per group (n)*	*RR (intervention vs control)*	*Required per group (n)*
1	0.83	587,000	0.83	587,000
2	0.78	94,000	0.69	48,000
3	0.69	25,000	0.65	19,000
4	0.76	27,000	0.64	11,000
5	0.80	29,000	0.67	10,000
6	0.83	34,000	0.71	10,000

RR: Relative risk.

Table 3. Relative risks of lung cancer death and required number per group by year for a 5-year risk of 3.5%.

Year	One annual screen		Three annual screens	
	RR (intervention vs control)	*Required per group (n)*	*RR (intervention vs control)*	*Required per group (n)*
1	0.83	368,000	0.83	368,000
2	0.78	63,000	0.69	35,000
3	0.69	16,000	0.65	13,000
4	0.76	20,000	0.64	8000
5	0.80	21,000	0.67	7000
6	0.83	24,000	0.71	7200

RR: Relative risk.

previously, there are good reasons why the population offered screening should be at higher risk than the general population. There are various ways of determining the risk criteria, and a lung cancer risk prediction model has an important role to play. Let us suppose that we plan to offer screening to a subgroup of the 55–64 years age group, determined by their absolute risk of lung cancer over 5 years. Of course, services may in the event, target a wider age group, possibly to the age of 70 or 75 years.

For example, one might specify a prevalence at first screen of 1%. The 2.06-year sojourn time and 96% sensitivity estimated by Chien and Chen imply that the prevalence at first screen will be approximately twice the annual incidence [15]. Thus, we would select a risk group with average lung cancer risk of 5 per thousand per year, that is, 2.5% over 5 years. As discussed previously, a minimum 5-year risk of 1.5% would give an average risk of 2.5%. Figure 1 shows the cumulative population frequency of cases and controls by individual absolute risk of lung cancer over 5 years in the age group 55–64 years from the LLP case–control study. Approximately 30% of controls and 70% of cases exceed this level of risk. Thus, if the eligibility criterion was set at a minimum 1.5% risk, this would mean that 30% of the population aged 55–64 years would qualify. For example, a woman with 35 years of smoking and a family history of lung cancer would qualify, whereas a woman with only 20 years of smoking and no family history would not. Since there are approximately 7 million men and women in this age group in the UK, this would mean that approximately 2.1 million people would be eligible for screening. In that 30% of the population aged 55–64 years, we would expect to find 70% of the lung cancer cases.

We can also use the risk estimates to choose a criterion based on the resources available. Suppose, for purposes of argument, that a CT scan costs approximately GB£150 per person, and that an annual budget of £100 million has been allocated for the screening. This approximately equates to screening 670,000 subjects per year, approximately 9% of the population aged 55–64 years. From Figure 1, approximately 9% of controls exceed 4% risk. We would, therefore, stipulate a minimum 5-year lung cancer risk of 4% to arrive at a screened population for which resources are available. For example, a man with 42 years smoking, a family history of lung cancer and a personal history of another malignancy would qualify, but a man with the family history and the personal history of another malignancy, but without the smoking history, would not. Approximately 35% of cases have an estimated risk exceeding this level, so in 9% of the population aged 55–64 years, we would expect to find 35% of the lung cancer cases. For a 2-yearly screening regime, we could double the eligible population. For 18% of the population to be eligible, the controls in Figure 1 indicate that a minimum 2.8% 5-year risk would be appropriate. This would capture approximately 50% of lung cancer cases.

Expert commentary

Tobacco control remains the most effective way to reduce morbidity and mortality from lung cancer. In much of the developed world, smoking rates have been declining for decades followed by corresponding reductions in lung-cancer risk. However, ex-smokers remain at high risk for years after stopping smoking and, in many countries, the majority of lung cancer cases occur in former rather than current smokers. Assuming further declines in smoking uptake, it will be important to identify nonsmokers and

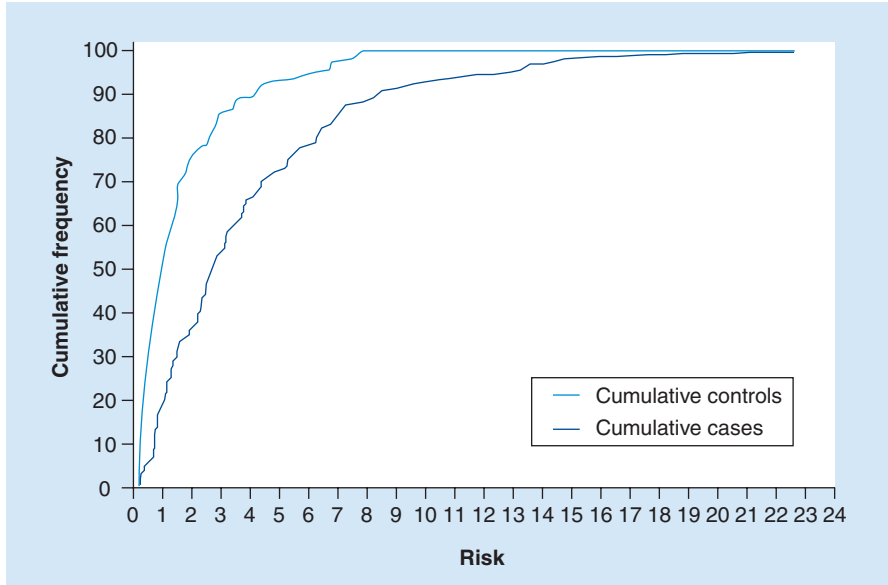

Figure 1. Cumulative frequencies of cases and controls aged 55–64 years, by 5-year percentage risk of lung cancer.

ex-smokers at relatively high risk of lung cancer. These observations have two implications: first, there is scope for further control of lung cancer mortality beyond tobacco control; and second, there is a need to identify those at increased risk of lung cancer for reasons other than smoking. Accurate delineation of risk of the disease through a combination of other risk factors with smoking will be important to identify subjects suitable for prevention or surveillance measures.

Five-year view

A possible means to reduce lung cancer mortality is by early detection with CT scanning. Several RCTs of CT imaging are ongoing in the USA and Europe to evaluate whether patient mortality can be improved by detecting and then treating lung cancer at an early stage. In the next 5 years, it is likely that new trials will be initiated and that some of the ongoing trials will report their results. The aforementioned work shows how the use of a simple risk model can contribute to the design of randomized trials to evaluate the technology and, in the event of positive results, to the design of a service screening program.

In terms of risk assessment, current research into genetic susceptibility and molecular biomarkers will lead to the discovery of genetic and epigenetic predictors of lung cancer risk [19]. The incorporation of these markers into risk models alongside epidemiological risk factors will be a major challenge.

Financial & competing interests disclosure

This work was partly supported by the NHS Health Technology Assessment programme and by the American Cancer Society. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Key issues

- Computed tomography (CT) screening for lung cancer is currently under evaluation.
- There is a need for accurate delineation of lung cancer risk to select populations suitable for CT screening in the event that the trials show that it is efficacious, or for other prevention or surveillance measures.
- As smoking uptake declines, nonsmoking risk factors take on a greater proportional role in risk prediction.
- A number of lung cancer risk prediction models are now available that combine smoking with other epidemiological factors, but in terms of simplicity, the Liverpool Lung Project (LLP) risk model is more directly applicable for use in the general population setting.
- Use of the LLP risk model has been demonstrated in assisting the design of randomized trials of lung cancer screening and in planning of services in the event of a positive result from the trials.

References

Papers of special note have been highlighted as:
- • of interest
- •• of considerable interest

1 Tyrer J, Duffy SW, Cuzick J. A breast cancer prediction model incorporating familial and personal risk factors. *Stat. Med.* 23, 1111–1130 (2004).

2 Park Y, Freeman AM, Gail MH *et al.* Validation of a colorectal cancer risk prediction model among white patients age 50 years and older. *J. Clin. Oncol.* 27, 694–698 (2009).

3 Neuhauser HK, Ellert U, Kurth BM. A comparison of Framingham and SCORE-based cardiovascular risk estimates in participants of the German National Health Interview and Examination Survey 1998. *Eur. J. Cardiovasc. Prev. Rehabil.* 12, 442–450 (2005).

4 Field JK, Duffy SW. Lung cancer screening: the way forward. *Br. J. Cancer* 99, 557–562 (2008).

• **Reviews the current information on computed tomography (CT) screening for lung cancer, including the need for assessing the impact of low-dose CT screening and treatment in decreasing lung cancer mortality compared with a control group, as well as accurate estimation of balance of benefits and harms, and determination of the cost–effectiveness of the intervention.**

5 van Klaveren RJ, de Koning HJ, Mulshine J, Hirsch FR. Lung cancer screening by spiral CT. What is the optimal target population for screening trials? *Lung Cancer* 38, 243–252 (2002).

•• **Provides recommendation for the selection of the optimal target population for lung cancer CT screening based on a combination of risk factors and high-risk populations.**

6 Bach PB, Elkin EB, Pastorino U *et al.* Benchmarking lung cancer mortality rates in current and former smokers. *Chest* 126, 1742–1749 (2004).

•• **Presents the development and validation of a risk model for lung cancer death.**

7 Spitz MR, Hong WK, Amos CI *et al.* A risk model for prediction of lung cancer. *J. Natl Cancer Inst.* 99, 715–726 (2007).

•• **Presents the development and validation of an absolute risk model for lung cancer.**

8 Cassidy A, Myles JP, van Tongeren M *et al.* The LLP risk model: an individual risk prediction model for lung cancer. *Br. J. Cancer* 98, 270–276 (2008).

•• **Presents the development and validation of a risk model for predicting a 5-year absolute risk of lung cancer for an individual with a specific combination of risk factors, including lifestyle and family history factors.**

9 Peto R. Influence of dose and duration of smoking on lung cancer rates. In: *Tobacco: A Major International Health Hazard*. Zaridze D, Peto R (Eds). International Agency for Research on Cancer, Lyon, France 23–33 (1986).

10 Cassidy A, Myles JP, Duffy SW, Liloglou T, Field JK. Family history and risk of lung cancer: age-at-diagnosis in cases and first degree relatives. *Br. J. Cancer* 95, 1288–1290 (2006).

11 Risch A, Plass C. Lung cancer epigenetics and genetics. *Int. J. Cancer* 123, 1–7 (2008).

12 Bach PB, Jett JR, Pastorino U, Tockman MS, Swensen SJ, Begg CB. Computed tomography screening and lung cancer outcomes. *JAMA* 297, 953–961 (2007).

13 Henschke CI, Yankelevitz D, Smith JP, Miettinen OS. Computed tomography screening for lung cancer. *JAMA* 298, 514–515 (2007).

14 Black WC. Computed tomography screening for lung cancer: review of screening principles and update on current status. *Cancer* 110, 2370–2384 (2007).

15 Chien CR, Chen THH. Mean sojourn time and effectiveness of mortality reduction for lung cancer screening with computed tomography. *Int. J. Cancer* 122, 2594–2599 (2008).

• **Provides current information on CT screening for lung cancer and presents mathematical models for projecting lung cancer mortality reduction attributed to screening.**

16 Coleman MP, Rachet B, Woods LM *et al.* Trends and socioeconomic inequalities in cancer survival in England and Wales up to 2001. *Br. J. Cancer* 90, 1367–1373 (2004).

17 Henschke CI, Yankelevitz DF, Libby DM, Pasmantier MW, Smith JP, Miettinen OS. Survival of patients with stage I lung cancer detected on CT screening. *N. Engl. J. Med.* 355, 1763–1771 (2006).

18 Marcus PM, Bergstralh EJ, Fagerstrom RM *et al*. Lung cancer mortality in the Mayo Lung Project: impact of extended follow-up. *J. Natl Cancer Inst.* 92, 1308–1316 (2000).

19 Field JK. Lung cancer risk models come of age. *Cancer Prev. Res.* 1, 226–228 (2008).

Website

101 UK Lung Cancer Screening Trial (UKLS) – Feasibility study and protocol development
www.hta.ac.uk/1752

Affiliations

• Stephen W Duffy, BSc, MSc, CStat
Professor of Cancer Screening, Cancer Research UK Centre for EMS, Wolfson Institute of Preventive Medicine, Barts and the London School of Medicine and Dentistry, Charterhouse Square, London EC1M 6BQ
Tel.: +44 207 882 3535
Fax: +44 207 882 0269
s.w.duffy@qmul.ac.uk

• Olaide Y Raji, PhD
Roy Castle Lung Cancer Research, School of Cancer Studies, The University of Liverpool Cancer Research Centre, 200 London Road, Liverpool, L3 9TA, UK
Tel.: +44 151 794 8889
rajio@roycastle.liv.ac.uk

• Olorunsola F Agbaje, PhD
Division of Cancer Studies, Cancer Epidemiology Unit, King's College London, School of Medicine, Academic Oncology, 3rd Floor, Bermondsey Wing, Guy's Hospital, London, SE1 9RT, UK
and
Academic Oncology, 3rd Floor, Bermondsey Wing, Guy's Hospital, London, SE1 9RT
Tel.: +44 020 7188 4862
olorunsola.agbaje@kcl.ac.uk

• Prue C Allgood
Cancer Research UK Centre for EMS, Wolfson Institute of Preventive Medicine, Barts and the London School of Medicine and Dentistry, Charterhouse Square, London, EC1M 6BQ, UK
Tel.: +44 207 882 3525
Fax: +44 207 882 3890
p.allgood@qmul.ac.uk

• Adrian Cassidy, PhD
Roy Castle Lung Cancer Research Programme, The University of Liverpool Cancer Research Centre, Liverpool, L3 9TA, UK
Tel.: +44 151 951 4793
Fax: +44 151 794 8989
cassidya@liv.ac.uk

• John K Field, MA, PhD, BDS, FRCPath, Professor, Roy Castle Lung Cancer Research Programme, The University of Liverpool Cancer Research Centre, 200 London Road, Liverpool, L3 9TA, UK
Tel.: +44 151 794 8900
j.k.field@liv.ac.uk