


PROTOCOL

Open Access



Development and validation of clinical prediction models to risk stratify patients presenting with small pulmonary nodules: a research protocol

Jason L. Oke^{1*} , Lyndsey C. Pickup⁵, Jérôme Declerck⁵, Matthew E. Callister³, David Baldwin⁴, Jennifer Gustafson², Heiko Peschl², Sarim Ather², Maria Tsakok², Alan Exell² and Fergus Gleeson²

Abstract

Introduction: Lung cancer is a common cancer, with over 1.3 million cases worldwide each year. Early diagnosis using computed tomography (CT) screening has been shown to reduce mortality but also detect non-malignant nodules that require follow-up scanning or alternative methods of investigation. Practical and accurate tools that can predict the probability that a lung nodule is benign or malignant will help reduce costs and the risk of morbidity and mortality associated with lung cancer.

Methods: Retrospectively collected data from 1500 patients with pulmonary nodule(s) of up to 15 mm detected on routinely performed CT chest scans aged 18 years old or older from three academic centres in the UK will be used to develop risk stratification models. Radiological, clinical and patient characteristics will be combined in multivariable logistic regression models to predict nodule malignancy. Data from over 1000 participants recruited in a prospective phase of the study will be used to evaluate model performance. Discrimination, calibration and clinical utility measures will be presented.

Keywords: Lung cancer, Pulmonary nodules, Clinical prediction model

Introduction

Small pulmonary nodules are a common finding on computed tomographic (CT) scans of the chest. Up to 75% of smokers scanned either as part of their clinical care or in lung cancer screening trials have sub-centimetre pulmonary nodules detected. The US National Lung Screening Trial (NLST) showed that up to 95% of lung nodules detected on CT scans of the chest were not malignant. The detection of nodules that are eventually proven to be benign may be expensive and both resource and time consuming with potential associated patient morbidity and mortality.

Current recommendations from internationally accepted Fleischner guidelines [1] and British Thoracic

Society (BTS) guidelines for the investigation and management of pulmonary nodules [2] suggests surveillance with CT for nodules of indeterminate risk (see Table 1).

A substantial proportion of pulmonary nodules detected on CT are judged to have an indeterminate risk of malignancy ($\approx 50\%$) but most ($\approx 97\%$) will be benign. Risk stratification tools that incorporate the age of the patient, their smoking history and their respiratory health could assist clinical decision making, reduce unnecessary investigations and quickly identify those at higher risk.

Existing nodule prediction models have been developed in highly selected patient groups with high rates of malignancy and give very different estimates of risk for smaller nodules. To date, five studies have derived composite prediction models based on a combination of clinical and radiological factors using multivariable logistic regression analysis [2]: Swensen (US, Mayo clinic model) [3], Gould (US, VA model) [4], Li (China) [5], Yonemori [6] (Japan)

*Correspondence: jason.oke@phc.ox.ac.uk

¹Nuffield Department of Primary Care Health Sciences, University of Oxford, Woodstock Road, OX2 6GG Oxford, UK

Full list of author information is available at the end of the article



Table 1 Abridged guideline recommendations following detection of incidentally detected lung nodules

Nodule diameter and volume	Recommended action
Fleischner guideline [1]	
< 6 mm or 100 mm ³	No routine follow-up if considered low risk or optional CT at 12 months if considered high risk (based on all relevant risk factors)
6–8mm or 100–250mm ³	Surveillance with CT
> 8 mm and 250 mm ³	CT, PET/CT or tissue sampling
British Thoracic Society guideline [2]	
< 5 mm or 80 mm ³	Discharge
5 – 8 mm or 80 – 300 mm ³	Surveillance with CT
> 8 mm and 300 mm ³	Assess using Brock model and surveillance with CT if risk < 10% or PET-CT if risk ≥ 10%

and McWilliams (Canada, Brock model) [7]. The study by Herder (Netherlands) et al. extended the Mayo clinic model to include positron emission tomography CT (PET-CT) results [8]. There are weaknesses in each of these predictive models particularly with respect to prediction of risk for small nodules, and their generalisability and evaluation. The Brock model was derived from a screening cohort of mostly current or former smokers. The VA model participants were mostly male smokers, and the Mayo cohort were a single cohort managed in the 1980s. The percentage of nodules that were malignant in the Yonemori, Li, VA and Mayo cohorts (75% [6], 62% [5], 54% [4] and 23% [3]) are unrepresentative of the risk in the wider population of patients with pulmonary nodules [9]. Validation of these models in a UK population extends to a single study in 263 patients identified from the lung cancer multidisciplinary team meeting and a nodule follow-up clinic between 2008 and 2013 [10]. Whilst these models discriminate well (C-statistic ranging from 73.5 to 91.2% [10]) they differ considerably in their estimate of risk for smaller nodules. For example, a 4 mm upper lobe nodule in a 70-year-old female smoker has a probability of malignancy of 0.3% according to the Brock model, 12% using the Mayo prediction model and 39% according the VA model [2].

This study aims to develop a clinical prediction model which will improve the accuracy of stratifying sub-centimetre lung nodules detected on chest CT scans. The study will incorporate solid nodules of 5 to 15 mm in diameter and aims improve the accuracy of stratifying lung nodules detected on chest CT scans across

a wide variety of scanner types, imaging protocols and patient populations. We hypothesise that further characterisation of sub-centimetre pulmonary nodules on chest CT scans will allow us, along with clinical profiling, to improve the accuracy of stratifying lung nodules as benign or malignant, and help guide their management. This will reduce the number of unnecessary investigations for benign nodules and may improve the ability to diagnose—or to instigate investigations that will diagnose malignant nodules earlier than is currently possible.

Methods

Sources of data

The development and validation of the risk stratification models for pulmonary nodules will be done using data from the artificial intelligence (AI) and Big Data for Early Lung Cancer Diagnosis (IDEAL) study. IDEAL is a National Institute for Health Research (NIHR) Invention for Innovation (i4i) funded project to build and evaluate a new computer-aided prediction model for malignancy in small pulmonary nodules.

Study design

The IDEAL study consists of two phases: a retrospective data collection phase and a prospective study. The risk stratification models will be derived using retrospectively collected data from the phase 1 of IDEAL and evaluated using prospective data collected in the phase 2. The model development team has access to clinical patient data, and to simple clinical observations about the CTs, made by radiologists. There is also an AI imaging model under independent development; this is trained using only data from CT images, with no patient data or human-derived CT observations.

Data extraction

CT chest scans reported as containing pulmonary nodules will be identified by a thoracic radiologist reporting the scan, or through an electronic search of CT chest scans previously performed on patients as part of their routine clinical care in the study sites. The scans will be anonymised prior to analysis.

Key study dates

The retrospective data collection phase began in January 2018 and is ongoing. The prospective study is expected to start in August 2018 and will end the completion of the last patient's follow-up (August 2020).

Participants

IDEAL is collecting data from three academic centres or *partners*; these are as follows:

- Oxford University Hospital NHS Foundation Trust
- Leeds Teaching Hospital NHS Trust
- Nottingham University Hospitals NHS Trust

Each of the three IDEAL partners is expected to contribute 500 patients to the phase 1 of the trial and at least 350 patients for phase 2 (see “[Sample size](#)” section for justification of sample size)

Inclusion and exclusion criteria

Inclusion criteria are the same for the phase 1 and phase 2 of IDEAL. A patient is eligible for inclusion in the study if they are as follows:

- Male or female, aged 18 years or above.
- Reported as having pulmonary nodule(s) of 5–15 mm detected on CT chest scan
- CT slice thickness of 3 mm or less.

The patient will *not* be included in the study if *any* of the following apply:

- Patient has more than 5 nodules of at least 5 mm.
- Technically inadequate CT scan (see [Appendix](#) for details).
- Diagnosis is unknown or could not be established.
- Current or prior history of malignancy in the last 5 years.

Outcome

The outcome or *ground truth* for each nodule will be established routinely in clinical care using the accepted published standards of the following:

- Histology
- 1 year for volume stability or 2 years for diameter stability, for benign nodules only
- Expert opinion, for subpleural or perifissural lymph nodes only
- Nodule resolution (i.e. infection clears up)

Benign nodules will be coded as zero, malignant nodules as 1.

Predictors

The following radiographic and clinical variables are available for inclusion into the risk stratification models. These have been selected because they either have been shown to be associated with the risk of nodule malignancy or benignity or have been used in other nodule prediction models.

We anticipate non-response to be an issue for variables such as “year when stopped smoking”, “smoking pack years” and “known industrial exposure” which may preclude their inclusion into the models (see the “[Handling missing data](#)” section for details of our method to handle missing data).

Sample size

A key concept in the consideration of the sample size for clinical prediction models for binary outcomes is the number of events per variable (EPV) [11]. The number of EPV is the number of events divided by the degrees of freedom considered in developing the prediction models. Roughly 10 EPVs have been proposed for accurate estimation of regression coefficients in a logistic regression model [12], whereas a minimum of 20 EPV are required to minimise differences between the bootstrap-corrected estimates and independent validation [11]. This implies that with a sample of 1500 at a prevalence of 10%, models with up to 15 degrees of freedom (df) can be accommodated and allow for accurate estimation of regression coefficients but may not ensure minimal differences between the performance of the models in the derivation and evaluation stage. Previous lung nodule risk models have required a small number of variables. For example, the full Brock model (2b) [7] used 12 df, Mayo clinic model [3] 7 df and VA model [4] 4 df.

Statistical analysis methods

We intend to build two clinical prediction models. A “full” model including all of the variables listed in Table 2 subject to the missing data criteria outlined previously, and a parsimonious model using a backwards selection criteria to drop variables that are not independently prognostic for malignancy.

Handling of predictors

We will assess whether any continuous predictors in the full-model exhibit a non-linear relationship with the risk of malignancy. In particular, we will carefully check for non-linearity of nodule size against risk, as the Brock model found it necessary to compensate for this [7].

Table 2 Candidate predictors

Radiographic	Clinical or patient characteristic
Number of nodules annotated	Age at CT scan
Does patient have other relevant nodules	Sex
Were there earlier scans reported negative	Smoking status
Nodule location	Number smoking pack years
Is nodule shape affected by emphysema	Year when stopped smoking
Maximum diameter measured on a 2d axial	Known COPD or emphysema
CT slice	Known coronary artery calcification
	Known pulmonary fibrosis
	Known industrial exposure
	Family history lung cancer

Non-linear variables will be modelled using fractional polynomials where appropriate [13].

Model-building procedure

Regression coefficients for both models will be estimated using maximum likelihood estimation in a logistic regression model. The open-source statistical software R [14] will be used with the *glm* function to estimate coefficients.

Handling missing data

For some of the clinical variables (such as family history of cancer), we anticipate missing data. To ensure generalisability and prevent loss of efficiency, we will explore methods for handling missing data either by creating a “missing” level for a factor or by using multiple imputation with chained equations (*mice*) [15]. This will ensure that we can utilise nodules with partially observed clinical data. We will only consider imputation for variables with less than 50% (across the whole cohort) of the data is missing, and we are confident that the missingness pattern can be considered “random” conditioning on the predictor variables in the models. Multiple imputation is considered valid under assumptions that the data are missing at random (MAR) dependent on the observed variables [16] but this assumption cannot be tested. The multiple imputation process will create m data sets (m to be determined later but likely to be between 10 and 50) and m models and m sets of parameter estimates. Parameter estimates for the final models will be combined using Rubin’s rules [17].

Internal validation

We will assess the *out of sample* performance of the models using bootstrap-based methods. This entails estimating the apparent performance of the models in the dataset used for development and then repeatedly drawing bootstrap samples (resampling with replacement) and re-estimating the models in order to obtain estimates of model-optimism. This is then subtracted from the apparent performance measure to obtain a optimism-correct performance measure [18].

Measures used to assess model performance

Discrimination will be summarised using the c-statistic (equivalent to the AUC) with 95% confidence intervals. Calibration of the models will be summarised by the intercept and slope of the validation curve. A calibration plot contrasting predicted probabilities with observed probabilities will be presented.

Identification of risk groups

Risk groups will be identified using false negative criteria of 0%, 0.5% and 1%. A false negative criteria has been selected because the prediction rule needs to have good *rule out* properties (high sensitivity). Using the predicted risk scores from the final model equations, we will

determine the thresholds which attain the desired false negative rate for the minimum number of false positives. As currently all indeterminate nodules are followed up, the true negatives rate represents the potential reduction in unnecessary investigations.

Prospective evaluation of the model

Missing data

We do not expect the same level of missing covariate data from the prospective data as in the retrospective data used for the model, and so, we will use only complete data for the validation process. If missing data exceeds 20%, then sensitivity analyses will be performed on a data set in which the missing covariate data set has been imputed as per the development stage.

Reporting

We will closely follow the TRIPOD guideline for transparent reporting of multivariable prediction models [19] and produce the following results from the model development stage.

- Diagram showing the flow of participants through the study, including the number of participants, with and without the outcome.
- Table of patient characteristics (demographic, clinical features and radiologic variables—including all candidate predictors in the models). Includes level of missing data per variable.
- Reporting of the unadjusted association between each candidate predictor and the outcome.
- Table of coefficient estimates (as beta coefficients and odds ratio) with confidence intervals.
- The prediction equation in full, with sufficient detail so that individual predictions can be made.
- Discrimination will be summarised using the c-statistic (equivalent to the AUC) with 95% confidence intervals.
- Intercept of validation curve (with perfect calibration the value would be 0).
- Slope of validation curve (with perfect calibration, the slope would be 1).
- A calibration plot
- False negative and false positive rates with 95% confidence intervals for the three decision thresholds defined in the model development stage.

Discussion

This protocol describes the methods and statistical analysis plan to develop and evaluate clinical prediction models for pulmonary nodules. Previous prediction models for lung nodules have been based on highly selected patient groups with high rates of malignancy and give very different estimates of risk for

smaller nodules. A robustly developed and validated clinical prediction model which generalises to a wide range of patients seen in clinical practice is highly desirable.

Appendix

Definition of technically inadequate CT

- 1 The series should have ≤ 3 mm slice thickness to be acceptable for markup. In practice, it is expected that scans with ≤ 2 mm are available and they are preferred to 3 mm thick scans.
- 2 The CT should not be from a tilted CT acquisition.
- 3 The CT should be free from artefacts which would affect the appearance of the nodule (motion) or the capacity of the clinician to make a diagnosis on the clinical image (noise). Such artefacts could manifest as follows:
 - Shifting structures in consecutive slices, which would be particularly visible in coronal or sagittal slices
 - Respiratory or cardiac motion
 - Excessive noise level that affects the nodule appearance.

Abbreviations

AI: Artificial intelligence; AUC: Area under the curve; BTS: British thoracic society; COPD: Chronic obstructive pulmonary disease; CT: Computed tomography; Df: Degrees of Freedom; EPV: Evens per variable; MAR: Missing at random; NIHR: National institute for health research; NHS: National health service; NLST: National lung screening trial; PET-CT: Positron emission tomography computed tomography; UK: United Kingdom; VA: Veterans affairs

Acknowledgements

Not applicable.

Funding

The work in this protocol is supported by an National Institute for Health Research, Invention for Innovation (i4i) award.

Availability of data and materials

The data that support the findings of this study are available from Oxford University Hospitals NHS Foundation Trust but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Any query regarding access to the data should be addressed to authors within the Oxford University Hospitals NHS Foundation Trust.

Authors' contributions

JLO, LCP and JD drafted the manuscript. All authors have read the manuscript and made critical revisions where appropriate. All authors read and approved the final manuscript.

Ethics approval and consent to participate

In keeping with the Governance Arrangements for Research Ethics Committees (GAfREC) and University Oxford Policy, research undertaken on data collected before formulation of a study, where data are anonymous to the researcher, does not require ethics approval. In keeping with the requirements of the Research Governance Framework, both sponsorship and Trust Management approval will be sought before the research is undertaken.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Nuffield Department of Primary Care Health Sciences, University of Oxford, Woodstock Road, OX2 6GG Oxford, UK. ²Oxford University Hospitals NHS Foundation Trust, Oxford, Oxford, UK. ³St James' University Hospital, Leeds, UK. ⁴Nottingham University Hospitals, Nottingham, UK. ⁵Optellum Ltd, Oxford, UK.

Received: 11 September 2018 Accepted: 13 November 2018

Published online: 29 November 2018

References

1. MacMahon H, Naidich DP, Goo JM, Lee KS, Leung ANC, Mayo JR, Mehta AC, Ohno Y, Powell CA, Prokop M, Rubin GD, Schaefer-Prokop CM, Travis WD, Van Schil PE, Bankier AA. Guidelines for management of incidental pulmonary nodules detected on CT images: from the Fleischner Society 2017. *Radiology*. 2017;284(1):228–43. <https://doi.org/10.1148/radiol.2017161659>.
2. Callister MEJ, Baldwin DR, Akram AR, Barnard S, Cane P, Draffan J, Franks K, Gleeson F, Graham R, Malhotra P, Prokop M, Rodger K, Subesinghe M, Waller D, Woolhouse I. British Thoracic Society guidelines for the investigation and management of pulmonary nodules: accredited by NICE. *Thorax*. 2015;70(Suppl 2):1–54. <https://doi.org/10.1136/thoraxjnl-2015-207168>.
3. Swensen SJ, Silverstein MD, Ilstrup DM, Schleck CD, Edell ES. The probability of malignancy in solitary pulmonary nodules: application to small radiologically indeterminate nodules. *Arch Intern Med*. 1997;157(8):849–55. <https://doi.org/10.1001/archinte.1997.00440290031002>.
4. Gould MK, Ananth L, Barnett PG. A clinical model to estimate the pretest probability of lung cancer in patients with solitary pulmonary nodules. *Chest*. 2007;131(2):383–8. <https://doi.org/10.1378/chest.06-1261>.
5. Li Y, Chen KZ, Wang J. Development and validation of a clinical prediction model to estimate the probability of malignancy in solitary pulmonary nodules in Chinese People. *Clin Lung Cancer*. 2011;12(5):313–9. <https://doi.org/10.1016/j.clcc.2011.06.005>.
6. Yonemori K, Tateishi U, Uno H, Yonemori Y, Tsuta K, Takeuchi M, Matsuno Y, Fujiwara Y, Asamura H, Kusumoto M. Development and validation of diagnostic prediction model for solitary pulmonary nodules. *Respirology*. 2007;12(6):856–62. <https://doi.org/10.1111/j.1440-1843.2007.01158.x>.
7. McWilliams A, Tammemagi MC, Mayo JR, Roberts H, Liu G, Soghrati K, Yasufuku K, Martel S, Laberge F, Gingras M, Atkar-Khattra S, Berg CD, Evans K, Finley R, Yee J, English J, Nasute P, Goffin J, Puka S, Stewart L, Tsai S, Johnston MR, Manos D, Nicholas G, Goss GD, Seely JM, Amjadi K, Tremblay A, Burrows P, MacEachern P, Bhatia R, Tsao MS, Lam S. Probability of cancer in pulmonary nodules detected on first screening CT. *N Engl J Med*. 2013;369(10):910–9. <https://doi.org/10.1056/NEJMoa1214726>.
8. Herder GJ, Van Tinteren H, Golding RP, Kostense PJ, Comans EF, Smit EF, Hoekstra OS. Clinical prediction model to characterize pulmonary nodules: validation and added value of 18F-fluorodeoxyglucose positron emission tomography. *Chest*. 2005;128(4):2490–6. <https://doi.org/10.1378/chest.128.4.2490>.
9. Callister MEJ, Baldwin DR. How should pulmonary nodules be optimally investigated and managed? *Lung Cancer*. 2016;91:48–55. <https://doi.org/10.1016/j.lungcan.2015.10.018>. <http://arxiv.org/abs/9503001>.
10. Al-Ameri A, Malhotra P, Thygesen H, Plant PK, Vaidyanathan S, Karthik S, Scarsbrook A, Callister MEJ. Risk of malignancy in pulmonary nodules: a validation study of four prediction models. *Lung Cancer*. 2015;89(1):27–30. <https://doi.org/10.1016/j.lungcan.2015.03.018>.
11. Austin PC, Steyerberg EW. Events per variable (EPV) and the relative performance of different strategies for estimating the out-of-sample validity of logistic regression models. *Stat Methods Med Res*. 2017;26(2):796–808. <https://doi.org/10.1177/0962280214558972>.
12. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis.

- J Clin Epidemiol. 1996;49(12):1373–9. [http://dx.doi.org/10.1016/S0895-4356\(96\)00236-3](http://dx.doi.org/10.1016/S0895-4356(96)00236-3).
13. Royston P, Ambler G, Sauerbrei W. The use of fractional polynomials to model continuous risk variables in epidemiology. *Int J Epidemiol*. 1999;28(5):964–74. <https://doi.org/10.1093/ije/28.5.964>.
 14. R Core Team R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2017. <https://www.R-project.org/>.
 15. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. *Stat Med*. 2011;30(4):377–99. <https://doi.org/10.1002/sim.4067>.
 16. Carpenter JR, Kenward MG. Multiple imputation and its application. Chichester: Wiley; 2013. pp. 75–89. <https://doi.org/10.1002/9781119942283.ch3>. <http://dx.doi.org/10.1002/9781119942283.ch3>.
 17. Marshall A, Altman DG, Holder RL, Royston P. Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines. *BMC Med Res Methodol*. 2009;9(1):1–8. <https://doi.org/10.1186/1471-2288-9-57>.
 18. Steyerberg EW. Clinical prediction models: a practical approach to development, validation, and updating. New York: Springer; 2009.
 19. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD Statement. *Eur Urol*. 2015;67(6):1142–51. <https://doi.org/10.1016/j.eururo.2014.11.025>.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

