



Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/186403/>

Version: Published Version

---

**Article:**

Gilbert, F.J., Harris, S., Miles, K.A. et al. (2022) Dynamic contrast-enhanced CT compared with positron emission tomography CT to characterise solitary pulmonary nodules : the SPUtNik diagnostic accuracy study and economic modelling. *Health Technology Assessment*, 26 (17). pp. 1-180. ISSN: 1366-5278

<https://doi.org/10.3310/wcei8321>

---

© Queen's Printer and Controller of HMSO 2022. This work was produced by Gilbert et al. under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK

**Reuse**

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.

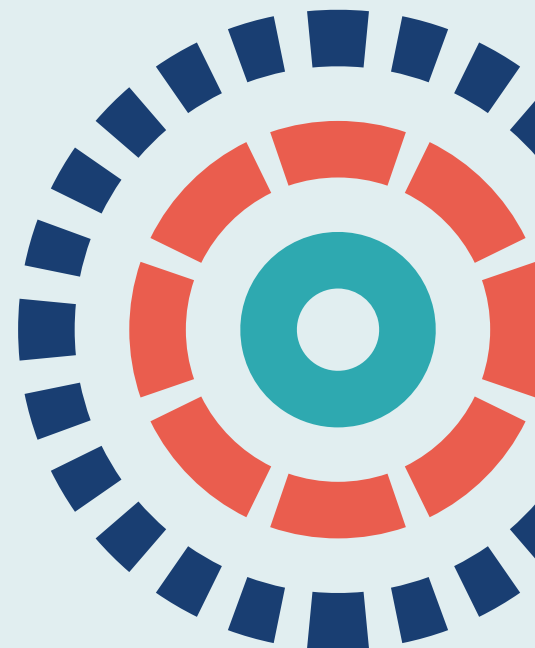
## Health Technology Assessment

Volume 26 • Issue 17 • March 2022

ISSN 1366-5278

# Dynamic contrast-enhanced CT compared with positron emission tomography CT to characterise solitary pulmonary nodules: the SPUtNik diagnostic accuracy study and economic modelling

*Fiona J Gilbert, Scott Harris, Kenneth A Miles, Jonathan R Weir-McCall, Nagmi R Qureshi, Robert C Rintoul, Sabina Dizdarevic, Lucy Pike, Donald Sinclair, Andrew Shah, Rosemary Eaton, Andrew Clegg, Valerio Benedetto, James E Hill, Andrew Cook, Dimitrios Tzelis, Luke Vale, Lucy Brindle, Jackie Madden, Kelly Cozens, Louisa A Little, Kathrin Eichhorst, Patricia Moate, Chris McClement, Charles Peebles, Anindo Banerjee, Sai Han, Fat Wui Poon, Ashley M Groves, Lutfi Kurban, Anthony J Frew, Matthew E Callister, Philip Crosbie, Fergus V Gleeson, Kavitasagary Karunasaagarar, Osei Kankam and Steve George on behalf of the SPUtNik investigators*





# Dynamic contrast-enhanced CT compared with positron emission tomography CT to characterise solitary pulmonary nodules: the SPUtNik diagnostic accuracy study and economic modelling

Fiona J Gilbert<sup>1\*</sup>, Scott Harris<sup>2</sup>, Kenneth A Miles<sup>1,3</sup>, Jonathan R Weir-McCall<sup>1</sup>, Nagmi R Qureshi<sup>3</sup>, Robert C Rintoul<sup>4,5</sup>, Sabina Dizdarevic<sup>6,7</sup>, Lucy Pike<sup>8</sup>, Donald Sinclair<sup>8</sup>, Andrew Shah<sup>9</sup>, Rosemary Eaton<sup>9</sup>, Andrew Clegg<sup>10</sup>, Valerio Benedetto<sup>10</sup>, James E Hill<sup>10</sup>, Andrew Cook<sup>11,12</sup>, Dimitrios Tzelis<sup>13</sup>, Luke Vale<sup>13</sup>, Lucy Brindle<sup>14</sup>, Jackie Madden<sup>11,12</sup>, Kelly Cozens<sup>11,12</sup>, Louisa A Little<sup>11,12</sup>, Kathrin Eichhorst<sup>11,12</sup>, Patricia Moate<sup>11,12†</sup>, Chris McClement<sup>11,12†</sup>, Charles Peebles<sup>15</sup>, Anindo Banerjee<sup>15</sup>, Sai Han<sup>16</sup>, Fat Wui Poon<sup>16</sup>, Ashley M Groves<sup>17</sup>, Lutfi Kurban<sup>18</sup>, Anthony J Frew<sup>6,7†</sup>, Matthew E Callister<sup>19</sup>, Philip Crosbie<sup>20</sup>, Fergus V Gleeson<sup>21,22</sup>, Kavitasagary Karunasaagar<sup>23</sup>, Osei Kankam<sup>24</sup> and Steve George<sup>11,12†</sup> on behalf of the SPUtNik investigators

<sup>1</sup>Department of Radiology, University of Cambridge School of Clinical Medicine, Biomedical Research Centre, University of Cambridge, Cambridge, UK

<sup>2</sup>Public Health Sciences and Medical Statistics, University of Southampton, Southampton, UK

<sup>3</sup>Department of Radiology, Royal Papworth Hospital, Cambridge, UK

<sup>4</sup>Department of Thoracic Oncology, Royal Papworth Hospital, Cambridge, UK

<sup>5</sup>Department of Oncology, University of Cambridge, Cambridge, UK

<sup>6</sup>Departments of Imaging and Nuclear Medicine and Respiratory Medicine, Brighton and Sussex University Hospitals NHS Trust, Brighton, UK

<sup>7</sup>Brighton and Sussex Medical School, Brighton, UK

<sup>8</sup>King's College London and Guy's and St Thomas' PET Centre, School of Biomedical Engineering and Imaging Sciences, King's College London, London, UK

<sup>9</sup>Radiation Protection Department, East and North Hertfordshire NHS Trust, Stevenage, UK

<sup>10</sup>Faculty of Health and Wellbeing, University of Central Lancashire, Preston, UK

<sup>11</sup>University Hospital Southampton NHS Foundation Trust, Southampton, UK

<sup>12</sup>Southampton Clinical Trials Unit, University of Southampton, Southampton, UK

<sup>13</sup>Population Health Sciences Institute, Newcastle University, Newcastle upon Tyne, UK

<sup>14</sup>School of Health Sciences, University of Southampton, Southampton, UK

<sup>15</sup>Department of Radiology and Respiratory Medicine, University Hospital Southampton NHS Foundation Trust, Southampton, UK

<sup>16</sup>West of Scotland PET Centre, Gartnavel Hospital, Glasgow, UK

<sup>17</sup>Institute of Nuclear Medicine, University College London, London, UK

<sup>18</sup>Department of Radiology, Aberdeen Royal Hospitals NHS Trust, Aberdeen, UK

<sup>19</sup>Department of Respiratory Medicine, Leeds Teaching Hospitals NHS Trust, Leeds, UK

<sup>20</sup>North West Lung Centre, University Hospital of South Manchester, Manchester, UK

<sup>21</sup>Department of Radiology, Churchill Hospital, Oxford, UK

<sup>22</sup>University of Oxford, Oxford, UK

<sup>23</sup>Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, UK

<sup>24</sup>Department of Thoracic Medicine, East Sussex Healthcare NHS Trust, Saint Leonards-on-Sea, UK

\*Corresponding author

†In memoriam

**Declared competing interests of authors:** Andrew Clegg, Valerio Benedetto and James E Hill are part-funded by the National Institute for Health Research (NIHR) Applied Research Collaboration North West Coast. Fiona J Gilbert is a NIHR Senior Investigator. She was a member of the Health Technology Assessment Efficacy and Mechanisms Evaluation Board (2014–19). Andrew Cook reports grants from the NIHR Health Technology Assessment programme during the conduct of the study. Fergus V Gleeson reports other from Optellum Ltd (Oxford, UK) outside the submitted work. Robert C Rintoul reports salary support from Cambridge Biomedical Research Centre and Cancer Research UK Cambridge Centre, during the conduct of the study. Robert C Rintoul also reports personal fees from AstraZeneca and Roche for work on their Advisory Board outside the submitted work. Nagmi R Qureshi is part-funded by the Cambridge Biomedical Research Centre.

Published March 2022

DOI: 10.3310/WCEI8321

This report should be referenced as follows:

Gilbert FJ, Harris S, Miles KA, Weir-McCall JR, Qureshi NR, Rintoul RC, *et al.* Dynamic contrast-enhanced CT compared with positron emission tomography CT to characterise solitary pulmonary nodules: the SPUtNik diagnostic accuracy study and economic modelling. *Health Technol Assess* 2022;**26**(17).

*Health Technology Assessment* is indexed and abstracted in *Index Medicus/MEDLINE*, *Excerpta Medica/EMBASE*, *Science Citation Index Expanded (SciSearch®)* and *Current Contents®/Clinical Medicine*.



# Health Technology Assessment

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Impact factor: 4.014

*Health Technology Assessment* is indexed in MEDLINE, CINAHL, EMBASE, the Cochrane Library and Clarivate Analytics Science Citation Index.

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) ([www.publicationethics.org/](http://www.publicationethics.org/)).

Editorial contact: [journals.library@nihr.ac.uk](mailto:journals.library@nihr.ac.uk)

The full HTA archive is freely available to view online at [www.journalslibrary.nihr.ac.uk/hta](http://www.journalslibrary.nihr.ac.uk/hta). Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: [www.journalslibrary.nihr.ac.uk](http://www.journalslibrary.nihr.ac.uk)

## Criteria for inclusion in the *Health Technology Assessment* journal

Reports are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

## HTA programme

Health Technology Assessment (HTA) research is undertaken where some evidence already exists to show that a technology can be effective and this needs to be compared to the current standard intervention to see which works best. Research can evaluate any intervention used in the treatment, prevention or diagnosis of disease, provided the study outcomes lead to findings that have the potential to be of direct benefit to NHS patients. Technologies in this context mean any method used to promote health; prevent and treat disease; and improve rehabilitation or long-term care. They are not confined to new drugs and include any intervention used in the treatment, prevention or diagnosis of disease.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

## This report

The research reported in this issue of the journal was funded by the HTA programme as project number 09/22/117. The contractual start date was in August 2012. The draft report began editorial review in December 2019 and was accepted for publication in September 2020. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

This report presents independent research funded by the National Institute for Health Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health and Social Care. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health and Social Care.

© Queen's Printer and Controller of HMSO 2022. This work was produced by Gilbert *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Published by the NIHR Journals Library ([www.journalslibrary.nihr.ac.uk](http://www.journalslibrary.nihr.ac.uk)), produced by Prepress Projects Ltd, Perth, Scotland ([www.prepress-projects.co.uk](http://www.prepress-projects.co.uk)).

## NIHR Journals Library Editor-in-Chief

---

**Professor Ken Stein** Professor of Public Health, University of Exeter Medical School, UK

## NIHR Journals Library Editors

---

**Professor John Powell** Chair of HTA and EME Editorial Board and Editor-in-Chief of HTA and EME journals. Consultant Clinical Adviser, National Institute for Health and Care Excellence (NICE), UK, and Professor of Digital Health Care, Nuffield Department of Primary Care Health Sciences, University of Oxford, UK

**Professor Andrée Le May** Chair of NIHR Journals Library Editorial Group (HSDR, PGfAR, PHR journals) and Editor-in-Chief of HSDR, PGfAR, PHR journals

**Professor Matthias Beck** Professor of Management, Cork University Business School, Department of Management and Marketing, University College Cork, Ireland

**Dr Tessa Crilly** Director, Crystal Blue Consulting Ltd, UK

**Dr Eugenia Cronin** Consultant in Public Health, Delta Public Health Consulting Ltd, UK

**Dr Peter Davidson** Consultant Advisor, Wessex Institute, University of Southampton, UK

**Ms Tara Lamont** Senior Adviser, Wessex Institute, University of Southampton, UK

**Dr Catriona McDaid** Reader in Trials, Department of Health Sciences, University of York, UK

**Professor William McGuire** Professor of Child Health, Hull York Medical School, University of York, UK

**Professor Geoffrey Meads** Emeritus Professor of Wellbeing Research, University of Winchester, UK

**Professor James Raftery** Professor of Health Technology Assessment, Wessex Institute, Faculty of Medicine, University of Southampton, UK

**Dr Rob Riemsma** Reviews Manager, Kleijnen Systematic Reviews Ltd, UK

**Professor Helen Roberts** Professor of Child Health Research, Child and Adolescent Mental Health, Palliative Care and Paediatrics Unit, Population Policy and Practice Programme, UCL Great Ormond Street Institute of Child Health, London, UK

**Professor Jonathan Ross** Professor of Sexual Health and HIV, University Hospital Birmingham, UK

**Professor Helen Snooks** Professor of Health Services Research, Institute of Life Science, College of Medicine, Swansea University, UK

**Professor Ken Stein** Professor of Public Health, University of Exeter Medical School, UK

**Professor Jim Thornton** Professor of Obstetrics and Gynaecology, Faculty of Medicine and Health Sciences, University of Nottingham, UK

Please visit the website for a list of editors: [www.journalslibrary.nihr.ac.uk/about/editors](http://www.journalslibrary.nihr.ac.uk/about/editors)

**Editorial contact:** [journals.library@nihr.ac.uk](mailto:journals.library@nihr.ac.uk)

# Abstract

## Dynamic contrast-enhanced CT compared with positron emission tomography CT to characterise solitary pulmonary nodules: the SPUtNik diagnostic accuracy study and economic modelling

Fiona J Gilbert<sup>1\*</sup>, Scott Harris<sup>2</sup>, Kenneth A Miles<sup>1,3</sup>, Jonathan R Weir-McCall<sup>1</sup>, Nagmi R Qureshi<sup>3</sup>, Robert C Rintoul<sup>4,5</sup>, Sabina Dizdarevic<sup>6,7</sup>, Lucy Pike<sup>8</sup>, Donald Sinclair<sup>8</sup>, Andrew Shah<sup>9</sup>, Rosemary Eaton<sup>9</sup>, Andrew Clegg<sup>10</sup>, Valerio Benedetto<sup>10</sup>, James E Hill<sup>10</sup>, Andrew Cook<sup>11,12</sup>, Dimitrios Tzelis<sup>13</sup>, Luke Vale<sup>13</sup>, Lucy Brindle<sup>14</sup>, Jackie Madden<sup>11,12</sup>, Kelly Cozens<sup>11,12</sup>, Louisa A Little<sup>11,12</sup>, Kathrin Eichhorst<sup>11,12</sup>, Patricia Moate<sup>11,12†</sup>, Chris McClement<sup>11,12†</sup>, Charles Peebles<sup>15</sup>, Anindo Banerjee<sup>15</sup>, Sai Han<sup>16</sup>, Fat Wui Poon<sup>16</sup>, Ashley M Groves<sup>17</sup>, Lutfi Kurban<sup>18</sup>, Anthony J Frew<sup>6,7†</sup>, Matthew E Callister<sup>19</sup>, Philip Crosbie<sup>20</sup>, Fergus V Gleeson<sup>21,22</sup>, Kavitasagary Karunasaagarar<sup>23</sup>, Osei Kankam<sup>24</sup> and Steve George<sup>11,12†</sup> on behalf of the SPUtNik investigators

<sup>1</sup>Department of Radiology, University of Cambridge School of Clinical Medicine, Biomedical Research Centre, University of Cambridge, Cambridge, UK

<sup>2</sup>Public Health Sciences and Medical Statistics, University of Southampton, Southampton, UK

<sup>3</sup>Department of Radiology, Royal Papworth Hospital, Cambridge, UK

<sup>4</sup>Department of Thoracic Oncology, Royal Papworth Hospital, Cambridge, UK

<sup>5</sup>Department of Oncology, University of Cambridge, Cambridge, UK

<sup>6</sup>Departments of Imaging and Nuclear Medicine and Respiratory Medicine, Brighton and Sussex University Hospitals NHS Trust, Brighton, UK

<sup>7</sup>Brighton and Sussex Medical School, Brighton, UK

<sup>8</sup>King's College London and Guy's and St Thomas' PET Centre, School of Biomedical Engineering and Imaging Sciences, King's College London, London, UK

<sup>9</sup>Radiation Protection Department, East and North Hertfordshire NHS Trust, Stevenage, UK

<sup>10</sup>Faculty of Health and Wellbeing, University of Central Lancashire, Preston, UK

<sup>11</sup>University Hospital Southampton NHS Foundation Trust, Southampton, UK

<sup>12</sup>Southampton Clinical Trials Unit, University of Southampton, Southampton, UK

<sup>13</sup>Population Health Sciences Institute, Newcastle University, Newcastle upon Tyne, UK

<sup>14</sup>School of Health Sciences, University of Southampton, Southampton, UK

<sup>15</sup>Department of Radiology and Respiratory Medicine, University Hospital Southampton NHS Foundation Trust, Southampton, UK

<sup>16</sup>West of Scotland PET Centre, Gartnavel Hospital, Glasgow, UK

<sup>17</sup>Institute of Nuclear Medicine, University College London, London, UK

<sup>18</sup>Department of Radiology, Aberdeen Royal Hospitals NHS Trust, Aberdeen, UK

<sup>19</sup>Department of Respiratory Medicine, Leeds Teaching Hospitals NHS Trust, Leeds, UK

<sup>20</sup>North West Lung Centre, University Hospital of South Manchester, Manchester, UK

<sup>21</sup>Department of Radiology, Churchill Hospital, Oxford, UK

<sup>22</sup>University of Oxford, Oxford, UK

<sup>23</sup>Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, UK

<sup>24</sup>Department of Thoracic Medicine, East Sussex Healthcare NHS Trust, Saint Leonards-on-Sea, UK

\*Corresponding author [fjg28@medschl.cam.ac.uk](mailto:fjg28@medschl.cam.ac.uk)

†In memoriam

**Background:** Current pathways recommend positron emission tomography–computerised tomography for the characterisation of solitary pulmonary nodules. Dynamic contrast-enhanced computerised tomography may be a more cost-effective approach.

**Objectives:** To determine the diagnostic performances of dynamic contrast-enhanced computerised tomography and positron emission tomography–computerised tomography in the NHS for solitary pulmonary nodules. Systematic reviews and a health economic evaluation contributed to the decision-analytic modelling to assess the likely costs and health outcomes resulting from incorporation of dynamic contrast-enhanced computerised tomography into management strategies.

**Design:** Multicentre comparative accuracy trial.

**Setting:** Secondary or tertiary outpatient settings at 16 hospitals in the UK.

**Participants:** Participants with solitary pulmonary nodules of  $\geq 8$  mm and of  $\leq 30$  mm in size with no malignancy in the previous 2 years were included.

**Interventions:** Baseline positron emission tomography–computerised tomography and dynamic contrast-enhanced computer tomography with 2 years' follow-up.

**Main outcome measures:** Primary outcome measures were sensitivity, specificity and diagnostic accuracy for positron emission tomography–computerised tomography and dynamic contrast-enhanced computerised tomography. Incremental cost-effectiveness ratios compared management strategies that used dynamic contrast-enhanced computerised tomography with management strategies that did not use dynamic contrast-enhanced computerised tomography.

**Results:** A total of 380 patients were recruited (median age 69 years). Of 312 patients with matched dynamic contrast-enhanced computer tomography and positron emission tomography–computerised tomography examinations, 191 (61%) were cancer patients. The sensitivity, specificity and diagnostic accuracy for positron emission tomography–computerised tomography and dynamic contrast-enhanced computer tomography were 72.8% (95% confidence interval 66.1% to 78.6%), 81.8% (95% confidence interval 74.0% to 87.7%), 76.3% (95% confidence interval 71.3% to 80.7%) and 95.3% (95% confidence interval 91.3% to 97.5%), 29.8% (95% confidence interval 22.3% to 38.4%) and 69.9% (95% confidence interval 64.6% to 74.7%), respectively. Exploratory modelling showed that maximum standardised uptake values had the best diagnostic accuracy, with an area under the curve of 0.87, which increased to 0.90 if combined with dynamic contrast-enhanced computerised tomography peak enhancement. The economic analysis showed that, over 24 months, dynamic contrast-enhanced computerised tomography was less costly (£3305, 95% confidence interval £2952 to £3746) than positron emission tomography–computerised tomography (£4013, 95% confidence interval £3673 to £4498) or a strategy combining the two tests (£4058, 95% confidence interval £3702 to £4547). Positron emission tomography–computerised tomography led to more patients with malignant nodules being correctly managed, 0.44 on average (95% confidence interval 0.39 to 0.49), compared with 0.40 (95% confidence interval 0.35 to 0.45); using both tests further increased this (0.47, 95% confidence interval 0.42 to 0.51).

**Limitations:** The high prevalence of malignancy in nodules observed in this trial, compared with that observed in nodules identified within screening programmes, limits the generalisation of the current results to nodules identified by screening.

**Conclusions:** Findings from this research indicate that positron emission tomography–computerised tomography is more accurate than dynamic contrast-enhanced computerised tomography for the characterisation of solitary pulmonary nodules. A combination of maximum standardised uptake value and peak enhancement had the highest accuracy with a small increase in costs. Findings from this research also indicate that a combined positron emission tomography–dynamic contrast-enhanced computerised tomography approach with a slightly higher willingness to pay to avoid missing small cancers or to avoid a ‘watch and wait’ policy may be an approach to consider.

**Future work:** Integration of the dynamic contrast-enhanced component into the positron emission tomography–computerised tomography examination and the feasibility of dynamic contrast-enhanced computerised tomography at lung screening for the characterisation of solitary pulmonary nodules should be explored, together with a lower radiation dose protocol.

**Study registration:** This study is registered as PROSPERO CRD42018112215 and CRD42019124299, and the trial is registered as ISRCTN30784948 and ClinicalTrials.gov NCT02013063.

**Funding:** This project was funded by the National Institute for Health Research (NIHR) Health Technology Assessment programme and will be published in full in *Health Technology Assessment*; Vol. 26, No. 17. See the NIHR Journals Library website for further project information.



# Contents

List of tables	xv
List of figures	xix
List of boxes	xxi
List of supplementary material	xxiii
Glossary	xxv
List of abbreviations	xxvii
Plain English summary	xxix
Scientific summary	xxx
<b>Chapter 1</b> Introduction	<b>1</b>
Background and rationale	1
<i>Imaging techniques</i>	1
<i>Therapeutic impact and cost-effectiveness of imaging for solitary pulmonary nodules</i>	2
<i>Incremental value of incorporating computerised tomography appearances of a solitary pulmonary nodule into the interpretation of integrated positron emission tomography-computerised tomography</i>	3
<i>Combined dynamic contrast-enhanced computerised tomography and positron emission tomography-computerised tomography</i>	3
<i>Incidental extrathoracic imaging findings</i>	3
<i>Rationale</i>	4
<i>Aims and objectives</i>	4
<b>Chapter 2</b> Observational study methods	<b>5</b>
Trial design	5
Participants	5
<i>Eligibility criteria</i>	5
<i>Changes to eligibility criteria after commencement of the trial</i>	6
Setting and recruitment pathway	6
Outcomes	7
<i>Primary outcome</i>	7
<i>Secondary outcomes</i>	8
Adverse events	8
Registration	8
Blinding	8
Data collection	8
<i>Baseline and evaluative procedures (visit 1, day 0)</i>	8
<i>Follow-up</i>	8
<i>Procedure when nodule was reduced in size or not visible on the dynamic contrast-enhanced computerised tomography scan</i>	9
<i>Withdrawal criteria</i>	9

## CONTENTS

Sample size	9
<i>Consideration for each test separately</i>	10
<i>Consideration for when both tests are used together</i>	10
Statistical methods and data analysis	10
Interim analysis and data storage	12
Accreditation of positron emission tomography centres	12
<i>Methods</i>	12
Accreditation of dynamic contrast-enhanced computerised tomography centres	12
Dynamic contrast-enhanced computerised tomography acquisition design	14
<b>Chapter 3 Site quality control</b>	<b>15</b>
Accreditation of positron emission tomography centres and technical quality control of positron emission tomography scans	15
Radiation dose	16
Radiographer-led quality control of dynamic contrast-enhanced computerised tomography scanners	16
Technical quality control of dynamic contrast-enhanced computerised tomography scans	17
<i>Deviation log</i>	17
Reconstructed field of view	17
Reconstructed slice thickness and slice interval	18
Contrast-enhanced imaging times	18
X-ray tube current and patient mass	18
Radiation dose	18
Reconstruction kernel	18
Dynamic contrast-enhanced computerised tomography injection rate	19
Window centre and width	19
Measurement software for contrast enhancement	19
Positron emission tomography and dynamic contrast-enhanced computerised tomography interobserver variability: site versus core read	19
<i>Positron emission tomography</i>	19
<i>Dynamic contrast-enhanced computerised tomography</i>	22
<b>Chapter 4 Diagnostic accuracy systematic review and meta-analysis</b>	<b>25</b>
Methods	25
Statistical analysis	26
Results	26
Discussion	29
<b>Chapter 5 Systematic review of cost-effectiveness</b>	<b>31</b>
Methods	31
<i>Search strategy</i>	31
<i>Study selection</i>	31
<i>Data extraction, critical appraisal and synthesis</i>	32
Results of systematic review of economic evaluations	32
<i>Study characteristics</i>	32
<i>Quality assessment</i>	36
<i>Discussion</i>	44
<i>Conclusion</i>	45
<b>Chapter 6 Main study results</b>	<b>47</b>
Centres	47
Screened patients	47
Recruited patients	47

Participant withdrawal	47
Participant follow-up	47
Numbers analysed	47
Baseline data and demographics	49
Baseline computerised tomography, fluorodeoxyglucose positron emission tomography-computerised tomography and dynamic contrast-enhanced computerised tomography results	51
Prevalence of cancer	57
Diagnostic accuracy of positron emission tomography-computerised tomography and dynamic contrast-enhanced computerised tomography	59
Combination of dynamic contrast-enhanced computerised tomography with positron emission tomography-computerised tomography	63
Assessing barriers to recruitment and actions taken to increase enrolment	67
<i>Overview</i>	67
<i>Clinical Research Network</i>	67
<i>Additional financial support at sites</i>	68
<i>Excess treatment costs/dynamic contrast-enhanced computerised tomography</i>	68
<i>Communication between NHS trusts</i>	68
<i>Overestimate of potential participants</i>	68
<i>Support of chest clinicians and radiologists</i>	68
<i>Support of multidisciplinary team co-ordinator</i>	68
<b>Chapter 7 Economic evaluation</b>	<b>69</b>
Introduction	69
Methods	69
<i>Model overview</i>	69
<i>Model parameters</i>	75
<i>Analysis</i>	81
<i>Model validation</i>	82
Results	82
<i>Base case</i>	82
<i>Sensitivity analyses</i>	85
Discussion	87
<b>Chapter 8 The IPCARD-SPN questionnaire substudy: the identification of patient-elicited symptoms that predict malignant pulmonary nodules – an exploratory analysis</b>	<b>89</b>
Background	89
Aims	89
Methods	89
<i>Recruitment and questionnaire completion</i>	89
<i>Data analysis</i>	89
Results	91
<i>Return and data completion rates</i>	91
<i>Covariates and interactions</i>	91
<i>Symptoms associated with lung cancer at year 1</i>	92
<i>Multivariable analyses</i>	92
<i>Model accuracy</i>	93
<i>Symptoms associated with lung cancer at year 2</i>	93
<i>Model accuracy</i>	94
Discussion	95
<i>Clinical plausibility and confounding</i>	95
<i>Limitations</i>	96
Conclusion	96

## CONTENTS

<b>Chapter 9 Discussion</b>	<b>97</b>
Main findings	97
<i>Study conduct</i>	97
<i>Study results</i>	97
<i>Health economics</i>	98
Strengths and limitations	99
Lessons for the future	100
Future research	100
<b>Chapter 10 Conclusion</b>	<b>103</b>
<b>Acknowledgements</b>	<b>105</b>
<b>References</b>	<b>109</b>
<b>Appendix 1 Patient and public involvement in the SPUtNIK study</b>	<b>119</b>
<b>Appendix 2 Methods tables</b>	<b>121</b>
<b>Appendix 3 The SPUtNIK trial: radiographer quality control measurements</b>	<b>123</b>
<b>Appendix 4 Number of scans that deviated from the scan protocol</b>	<b>125</b>
<b>Appendix 5 Diagnostic accuracy: additional tables</b>	<b>131</b>
<b>Appendix 6 MEDLINE search strategy</b>	<b>139</b>
<b>Appendix 7 Critical appraisal checklist</b>	<b>141</b>
<b>Appendix 8 List of excluded studies</b>	<b>143</b>
<b>Appendix 9 Systematic review of economic evaluations: tables</b>	<b>145</b>
<b>Appendix 10 List of centres and principal investigators</b>	<b>153</b>
<b>Appendix 11 Main results: additional tables</b>	<b>155</b>
<b>Appendix 12 Applying likelihood ratios to update prevalence of disease calculations</b>	<b>161</b>
<b>Appendix 13 Parameter tables</b>	<b>163</b>
<b>Appendix 14 Diagnostic yield and accuracy of lung biopsy</b>	<b>167</b>
<b>Appendix 15 Biopsy and surgery complications</b>	<b>171</b>
<b>Appendix 16 Data used in sensitivity analyses</b>	<b>177</b>
<b>Appendix 17 The IPCARD-SPN questionnaire substudy: additional tables</b>	<b>179</b>

# List of tables

<b>TABLE 1</b> Schedule of observations and procedures	<b>7</b>
<b>TABLE 2</b> Technical checks performed on the PET data	<b>15</b>
<b>TABLE 3</b> Semiquantitative grading scheme for both the PET and CT component of the PET/CT examination combining the anatomical features from the CT and the metabolic features from the PET exam	<b>20</b>
<b>TABLE 4</b> Inclusion criteria for systematic review of economic evaluations	<b>31</b>
<b>TABLE 5</b> Characteristics of included studies	<b>34</b>
<b>TABLE 6</b> Summary of model inputs in the included studies	<b>38</b>
<b>TABLE 7</b> Baseline characteristics and medical history of the study participants	<b>49</b>
<b>TABLE 8</b> The CT, DCE-CT and PET/CT information	<b>51</b>
<b>TABLE 9</b> The SPN sizes according to each imaging technique and split by cancer status ( $N = 312$ )	<b>55</b>
<b>TABLE 10</b> Confirmation of malignant SPN following 2 years' follow-up ( $N = 312$ )	<b>57</b>
<b>TABLE 11</b> Final nodule diagnosis by smoking status ( $N = 312$ )	<b>58</b>
<b>TABLE 12</b> Extrathoracic findings from PET/CT ( $N = 312$ )	<b>58</b>
<b>TABLE 13</b> Grading of SPN and nodule growth by final outcome status ( $N = 312$ )	<b>59</b>
<b>TABLE 14</b> Diagnostic performance of each imaging technique: predefined thresholds ( $N = 312$ )	<b>62</b>
<b>TABLE 15</b> Diagnostic performance of the prespecified and best-performing exploratory models ( $N = 306$ )	<b>64</b>
<b>TABLE 16</b> Diagnostic performance of the prespecified and best-performing exploratory models ( $N = 312$ )	<b>65</b>
<b>TABLE 17</b> Classification of results based on test outcome and final nodule status	<b>71</b>
<b>TABLE 18</b> Summary of outcome values at each terminal node	<b>79</b>
<b>TABLE 19</b> Costs and consequences results for base-case analysis	<b>83</b>
<b>TABLE 20</b> Cost-effectiveness results for base case analysis	<b>84</b>
<b>TABLE 21</b> Costs and consequences results of scenario excluding indeterminate results	<b>86</b>

<b>TABLE 22</b> Odds ratios and <i>p</i> -values for symptoms associated with lung cancer at year 1 ( $p \leq 0.05$ ), adjusted for age and sex	92
<b>TABLE 23</b> Multivariable models for lung cancer diagnosis at year 1 (model 1: entry of variables significant at 0.05)	92
<b>TABLE 24</b> Independent associations with lung cancer for variables re-entered into model 1	94
<b>TABLE 25</b> The ORs and <i>p</i> -values for symptoms associated with lung cancer at year 2 ( $p \leq 0.05$ ), adjusted for age and sex	94
<b>TABLE 26</b> Multivariable model for lung cancer diagnosis at year 2	94
<b>TABLE 27</b> List of recruiting sites and PET centres	121
<b>TABLE 28</b> List of scanners used by site	122
<b>TABLE 29</b> Radiographer phantom QC protocol	123
<b>TABLE 30</b> Number of scans that deviated from the scan protocol	125
<b>TABLE 31</b> Summary of the study design and baseline characteristics of those included in the meta-analysis	132
<b>TABLE 32</b> Summary of the CT acquisition protocols and measurements of the studies included in the meta-analysis	133
<b>TABLE 33</b> The QUADAS-2 components for each of the individual studies	135
<b>TABLE 34</b> Diagnostic performance of DCE-CT for the evaluation of pulmonary nodules	135
<b>TABLE 35</b> Subgroup analyses of the diagnostic performance of DCE-CT for the evaluation of indeterminate pulmonary lesions	136
<b>TABLE 36</b> Diagnostic pathways evaluated in included studies	145
<b>TABLE 37</b> Quality assessment of included studies	146
<b>TABLE 38</b> Sensitivity analysis reported in included papers	148
<b>TABLE 39</b> Cost-effectiveness of individual tests and diagnostic strategies for SPNs	150
<b>TABLE 40</b> The ICERs for non-dominated strategies at different pre-test probabilities of malignancy reported by Gould <i>et al.</i>	152
<b>TABLE 41</b> Trial recruitment and data availability summary by site	155
<b>TABLE 42</b> Diagnostic performance of each imaging technique: pre-defined thresholds with a 50% prevalence ( $N = 312$ )	156
<b>TABLE 43</b> Diagnostic performance of each imaging technique: pre-defined thresholds with a 70% prevalence ( $N = 312$ )	156

<b>TABLE 44</b> Diagnostic performance of each imaging technique: pre-defined thresholds in subset of patients with a SPN of 8–15 mm on initial CT scan ( $N = 167$ )	<b>157</b>
<b>TABLE 45</b> Diagnostic performance of each imaging technique: pre-defined thresholds in subset of patients with a SPN of $> 15$ mm to $\leq 20$ mm on initial CT scan ( $N = 74$ )	<b>157</b>
<b>TABLE 46</b> Diagnostic performance of each imaging technique: pre-defined thresholds in subset of patients with a SPN of $> 20$ mm on initial CT scan ( $N = 65$ )	<b>158</b>
<b>TABLE 47</b> Number of participants with an adverse event who reported each type	<b>159</b>
<b>TABLE 48</b> Notation of diagnostic results of $^{18}\text{F}$ -FDG-PET/CT according to malignancy status	<b>161</b>
<b>TABLE 49</b> Diagnostic results of $^{18}\text{F}$ -FDG-PET/CT according to malignancy status	<b>161</b>
<b>TABLE 50</b> Clinical parameters used in the model	<b>163</b>
<b>TABLE 51</b> Estimated proportions for procedure-related non-fatal complications and mortality	<b>163</b>
<b>TABLE 52</b> Distribution of management options following initial imaging test results (from trial)	<b>164</b>
<b>TABLE 53</b> Unit costs of imaging tests and treatment procedures	<b>165</b>
<b>TABLE 54</b> Life expectancy according to malignancy and treatment status	<b>166</b>
<b>TABLE 55</b> Studies reported in the BTS guidelines, included in the meta-analysis	<b>169</b>
<b>TABLE 56</b> Inclusion/exclusion criteria and search terms	<b>171</b>
<b>TABLE 57</b> Summary of biopsy complications reported in identified sources	<b>171</b>
<b>TABLE 58</b> Biopsy complication rates used in published economic evaluations (see <i>Chapter 6</i> )	<b>172</b>
<b>TABLE 59</b> Summary of lobectomy and wedge-resection procedure-related complications reported in identified sources	<b>173</b>
<b>TABLE 60</b> Surgical complication and mortality rates in published economic evaluations	<b>174</b>
<b>TABLE 61</b> Search strategy for radiotherapy complications	<b>175</b>
<b>TABLE 62</b> Imaging test accuracy for scenario analysis	<b>177</b>
<b>TABLE 63</b> Baseline characteristics for the SPUtNik study subsample with valid IPCARD-SPN questionnaire data and valid lung cancer outcome data	<b>179</b>
<b>TABLE 64</b> Predicted probabilities for the diagnosis of lung cancer within the first 12 months of surveillance (model 1)	<b>180</b>



# List of figures

<b>FIGURE 1</b> Flow diagram of the SPUtNik trial	6
<b>FIGURE 2</b> The SPUtNik chest-equivalent phantom	13
<b>FIGURE 3</b> The SPUtNik water-filled radiographer quality assurance phantom	13
<b>FIGURE 4</b> Injected activity as a function of patient's weight for all patients	16
<b>FIGURE 5</b> Uptake time (time between injection and PET scan start) for all patients	16
<b>FIGURE 6</b> Bland–Altman plot comparing the SUVmax obtained at the recruiting site with that obtained at the core laboratory	21
<b>FIGURE 7</b> Bland–Altman plot comparing maximum enhancement obtained at the recruiting site with that obtained at the core laboratory	23
<b>FIGURE 8</b> The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram of the papers identified by the literature search, screened for eligibility and included in the final study	27
<b>FIGURE 9</b> The QUADAS-2 scoring summary of the included studies	28
<b>FIGURE 10</b> Bivariate SROC curve of the included studies	28
<b>FIGURE 11</b> Study selection: original and updated searches	33
<b>FIGURE 12</b> The SPUtNik study standards for the reporting of diagnostic accuracy studies (STARD) flow diagram	48
<b>FIGURE 13</b> Sensitivity, specificity and diagnostic accuracy by nodule size: PET/CT and DCE-CT	61
<b>FIGURE 14</b> Sensitivity, specificity and diagnostic accuracy by nodule size: PET/CT and SUVmax > 2.5	61
<b>FIGURE 15</b> The ROC curves for the diagnostic performance of the prespecified and best-performing exploratory models (N = 300)	63
<b>FIGURE 16</b> Simplified scheme of decision tree, including clones of subtrees	73
<b>FIGURE 17</b> The cost-effectiveness acceptability curve for cost per correctly treated malignancy	84
<b>FIGURE 18</b> The cost-effectiveness acceptability curve for cost per correctly managed case	84
<b>FIGURE 19</b> Tornado diagram showing impact on incremental costs of changes in key parameters	85

## LIST OF FIGURES

<b>FIGURE 20</b> Tornado diagram showing the impact on the incremental cost per correctly managed case	<b>86</b>
<b>FIGURE 21</b> The ROC curve for model 1	<b>93</b>
<b>FIGURE 22</b> The ROC curve for model 2	<b>95</b>
<b>FIGURE 23</b> Measurement positions in QC phantom	<b>124</b>
<b>FIGURE 24</b> Forest plot of the included studies	<b>137</b>
<b>FIGURE 25</b> Galbraith plot examining interstudy heterogeneity for publication bias by incorporating the effect size of each study, compared with the pooled analysis	<b>137</b>
<b>FIGURE 26</b> The SROC curve of biopsy diagnostic accuracy	<b>169</b>

# List of boxes

**BOX 1** Summary of outcomes included in the decision tree

74



# List of supplementary material

**Report Supplementary Material 1** Determination of correction factors for incorrect injection rate during DCE-CT examinations

Supplementary material can be found on the NIHR Journals Library report page (<https://doi.org/10.3310/WCEI8321>).

Supplementary material has been provided by the authors to support the report and any files provided at submission will have been seen by peer reviewers, but not extensively reviewed. Any supplementary material provided at a later stage in the process may not have been peer reviewed.



# Glossary

**Chief investigator** The person who takes overall responsibility for the conduct of research.

**Consolidated Standards of Reporting Trials diagram** A diagram that shows the flow of patients through a clinical trial ([www.consort-statement.org/consort-statement/flow-diagram](http://www.consort-statement.org/consort-statement/flow-diagram); accessed 25 September 2018).

**Electronic case report form** A way of collecting data on patients in clinical trials.

**Intermediate profession** Part of the UK National Statistics Socio-economic Classification. These are occupations described as involving clerical work, sales and service.

**Principal investigator** The lead investigator at a clinical site.

**Qualitative** As it pertains to research, a broad selection of research methods, aimed at exploring the how and why of a process, not just the countable outcomes.

**Quality-adjusted life-year** A way of accounting for both the duration and quality of an outcome.

**Quantitative** As it pertains to research, the investigation of a research question using statistical, mathematical or computational techniques.

**Within trial** All the data used for analysis are collected during the trial. No estimation is made of what may happen once the trial has finished.



## List of abbreviations

(alt)	second alternative approach	NICE	National Institute for Health and Care Excellence
<sup>18</sup> F-FDG	fluorine-18-labelled-fluorodeoxyglucose	NIHR	National Institute for Health Research
AIC	Akaike information criterion	NLR	negative likelihood ratio
AUROC	area under receiver operating characteristic	NPV	negative predictive value
BTS	British Thoracic Society	ODA	overall diagnostic accuracy
CCA	cost-consequences analysis	OR	odds ratio
CCG	Clinical Commissioning Group	PET	positron emission tomography
CI	confidence interval	PET/CT	positron emission tomography-computerised tomography
COPD	chronic obstructive pulmonary disease	PI	principal investigator
CRF	case report form	PPI	patient and public involvement
CRN	Clinical Research Network	PLR	positive likelihood ratio
CT	computerised tomography	PPV	positive predictive value
DCE-CT	dynamic contrast-enhanced computerised tomography	PSF	point-spread function
DOR	diagnostic odds ratio	QA	quality assurance
DRL	diagnostic reference level	QALY	quality-adjusted life-year
FDG	fluorodeoxyglucose	QC	quality control
HTA	Health Technology Assessment	QECT	quantitative contrast-enhanced computerised tomography
HU	Hounsfield unit	QUADAS-2	Quality Assessment of Diagnostic Accuracy Studies
ICER	incremental cost-effectiveness ratio	R&D	research and development
IPCARD	Identifying symptoms that Predict Chest And Respiratory Disease	ROC	receiver operating characteristic
IPCARD-SPN	Identifying symptoms that Predict Chest And Respiratory Disease – Solitary Pulmonary Nodule	ROI	region of interest
LOA	limit of agreement	SABR	stereotactic ablative radiotherapy
MDT	multidisciplinary team	SCTU	Southampton Clinical Trials Unit
MRI	magnetic resonance imaging	SPN	solitary pulmonary nodule
		SPUtNik	Single PULmonary Nodule Investigation

## LIST OF ABBREVIATIONS

SPV	standardised perfusion value	SUVmean	mean standardised uptake value
SROC	summary receiver operating characteristic	TMG	Trial Management Group
SUV	standardised uptake value	TNB	transthoracic needle biopsy
SUVmax	maximum standardised uptake value	TSC	Trial Steering Committee
		WTP	willingness to pay

## Plain English summary

A nodule found on a lung scan can cause concern as it may be a sign of cancer. Finding lung cancer nodules when they are small (i.e. < 3 cm) is very important. Most nodules are not cancerous. Computerised tomography (cross-sectional images created from multiple X-rays) and positron emission tomography-computerised tomography (a technique that uses a radioactive tracer combined with computerised tomography) are used to see whether or not a nodule is cancerous; although they perform well, improvements are required.

This study compared dynamic contrast-enhanced computerised tomography with positron emission tomography-computerised tomography scans to find out which test is best. Dynamic contrast-enhanced computerised tomography involves injection of a special dye into the bloodstream, followed by repeated scans of the nodule over several minutes. We assessed the costs to the NHS of undertaking the different scans, relative to their benefits, to judge which option was the best value for money.

We recruited 380 patients from 16 hospitals across England and Scotland, of whom 312 had both dynamic contrast-enhanced computerised tomography and positron emission tomography-computerised tomography scans. We found that current positron emission tomography-computerised tomography is more accurate, providing a correct diagnosis in 76% of cases, than the new dynamic contrast-enhanced computerised tomography, which provides a correct diagnosis in 70% of cases. Although dynamic contrast-enhanced computerised tomography cannot replace positron emission tomography-computerised tomography, it may represent good-value use of NHS resources, especially if it is performed before positron emission tomography-computerised tomography and they are used in combination.

Although more research is required, it may be possible in the future to perform dynamic contrast-enhanced computerised tomography at the same time as positron emission tomography-computerised tomography in patients with suspected lung cancer or if a lung nodule is found on a lung screening programme at the time of the computerised tomography examination. This may reduce the need for some people to have positron emission tomography-computerised tomography.



# Scientific summary

## Background

Patients presenting with a solitary pulmonary nodule on diagnostic imaging tests are an important group because if early-stage lung cancer is treated with curative intent with surgical resection, the 5-year survival rate approaches 70%. However, not all solitary pulmonary nodules are due to lung cancer; the accurate characterisation of solitary pulmonary nodules is a diagnostic challenge with significant resource and cost implications. Owing to the association between nodule size and likelihood of malignancy, current management strategies are directed by nodule size. Nodules of < 5 mm require no follow-up, whereas nodules of > 8 mm in diameter require further diagnostic workup with positron emission tomography–computerised tomography or biopsy. However, both approaches are expensive, and positron emission tomography–computerised tomography is not available at all hospitals in the UK. Dynamic contrast-enhanced computerised tomography quantifies the degree of enhancement of pulmonary nodules following the intravenous administration of iodine-based contrast. The degree of enhancement reflects the extent of vascularity of the nodule, which has been shown to be an accurate marker for the diagnosis of solitary pulmonary nodules, with high sensitivity and moderate specificity. Moreover, dynamic contrast-enhanced computerised tomography has been shown to be a more cost-effective approach in the diagnostic workup of nodules than positron emission tomography–computerised tomography. However, the studies underpinning this analysis are weak. They predominantly involved single-centre studies, did not directly compare positron emission tomography with dynamic contrast-enhanced computerised tomography, included pulmonary masses as well as nodules, and incorporated studies using old technology whereby the spatial resolution of positron emission tomography–computerised tomography did not allow for accurate assessment of smaller nodules. Furthermore, these technologies have not been compared for their impact and cost-effectiveness within the NHS structure.

The aim of the current study was to address these gaps in knowledge through the conduct of a large multicentre comparative accuracy trial of dynamic contrast-enhanced computerised tomography with positron emission tomography–computerised tomography in the assessment of solitary pulmonary nodules. Prior to commencing the trial, two systematic reviews were conducted examining diagnostic accuracy studies and economic evaluations of alternative diagnostic imaging techniques (whether individually or in combination) for characterising solitary pulmonary nodules (PROSPERO registration numbers CRD42018112215 and CRD42019124299, respectively). In addition to assessing the different techniques, the systematic reviews provided data relevant to the NHS for decision-analytic modelling.

## Objectives

The primary objectives were to:

- determine, with high precision, the diagnostic performances of dynamic contrast-enhanced computerised tomography and fluorine-18-labelled-fluorodeoxyglucose positron emission tomography–computerised tomography in the NHS for the characterisation of solitary pulmonary nodules
- use decision-analytic modelling to assess the likely costs and health outcomes resulting from incorporation of dynamic contrast-enhanced computerised tomography into management strategies for patients with solitary pulmonary nodules.

The secondary objectives were to:

- assess, in an NHS setting, the incremental value of incorporating the computerised tomography appearances of a solitary pulmonary nodule into the interpretation of integrated positron emission tomography–computerised tomography examinations
- assess whether or not combining dynamic contrast-enhanced computerised tomography with positron emission tomography–computerised tomography is more accurate and/or cost-effective in the characterisation of solitary pulmonary nodules than either test used alone or in series
- document the nature and incidence of incidental extrathoracic findings on positron emission tomography–computerised tomography and dynamic contrast-enhanced computerised tomography undertaken for the characterisation of solitary pulmonary nodules and to model the cost-effectiveness of the two techniques.

## Methods

### *Systematic review of economic evaluation*

Studies were identified through searches of six electronic databases, from inception to November 2018 [i.e. MEDLINE, EMBASE, Web of Science, Bioscience Information Service (BIOSIS), The Cochrane Library and Science Direct]; through reference lists of relevant studies; and by contacting experts. Cost, cost-effectiveness and cost–utility studies were included that compared strategies involving computerised tomography or positron emission tomography with resection, biopsy and/or clinical follow-up. Studies had to report either costs, cost per case detected or incremental cost per life-year/quality-adjusted life-year gained. Studies were selected through two stages: initially, titles and abstracts were screened, with manuscripts of selected papers then retrieved for full-text screening. At both stages, two reviewers independently screened studies. Data were extracted into a standard template by one reviewer and checked by a second reviewer. Differences at all stages were resolved through discussion. Data were synthesised through narrative review and studies were critically appraised using standard criteria.

### *Multicentre comparative accuracy trial*

The trial was designed in accordance with the guidance for the methods of technology appraisal issued by the National Institute for Health and Care Excellence and adopted by the National Institute for Health and Care Excellence in formulating its guidance for the use of positron emission tomography in the staging of lung cancer.

The trial was designed as a prospective, multicentre observational study to assess the diagnostic performance and incremental value of dynamic contrast-enhanced computerised tomography by addition of this modality to positron emission tomography–computerised tomography in a cohort of 375 patients with solitary pulmonary nodules. Preceding symptoms were collected in a questionnaire at recruitment and at 1 year, and were correlated with the outcome.

### *Settings and participants*

Participants with a solitary pulmonary nodule were recruited from secondary or tertiary outpatient settings at 16 hospitals in the UK.

### *Inclusion criteria*

- A soft-tissue solitary dominant pulmonary nodule of  $\geq 8$  mm and  $\leq 30$  mm on the axial plane:
  - measured on lung window using conventional computerised tomography
  - no other ancillary evidence strongly indicative of malignancy (e.g. distant metastases or unequivocal local invasion)
  - if clinicians and reporting radiologists believed that the patient was being treated as having a single pulmonary nodule and there were other small lesions of  $< 4$  mm that would normally be disregarded, the patient was included in the trial.

- Nodules already under surveillance were included provided that the patient had recently undergone, or had scheduled, positron emission tomography–computerised tomography.
- Aged  $\geq 18$  years at the time of providing consent.
- Able and willing to consent to the trial.

### Exclusion criteria

- Pregnancy.
- History of malignancy in the previous 2 years.
- Confirmed aetiology of the nodule at the time of the qualifying computerised tomography scan; as this was a diagnostic study, if the aetiology of the nodule was confirmed by investigation such as positron emission tomography–computerised tomography or bronchoscopy prior to consent, the patient remained eligible, as the intention to include was made on the analysis of the qualifying computerised tomography scan.
- Biopsy of nodule prior to dynamic contrast-enhanced computerised tomography.
- Contraindication to potential radiotherapy or surgery.
- Contraindication to imaging techniques (assessed by local procedures).

### Outcomes

The primary outcome was the diagnostic accuracy of dynamic contrast-enhanced computerised tomography and positron emission tomography–computerised tomography, and their cost-effectiveness in the diagnostic workup of solitary pulmonary nodules.

The secondary outcomes were as follows:

- the accuracy (sensitivity, specificity, overall diagnostic accuracy) of the full spectrum of cut-off points, alone or in combination, of the two tests
- the incremental cost per malignant case treated and the incremental cost per correctly managed case
- the presence of incidental extrathoracic lesions requiring investigation.

The reference standard was histology and/or completion of 2 years of nodule surveillance. For a nodule to be diagnosed as malignant, histological confirmation was required, or an increase in nodule size with multidisciplinary team certainty of malignancy when biopsy/resection was not possible. Benign status was established through either histology or demonstration of stability over 2 years of computerised tomography monitoring. A biopsy of a lesion increasing in size during follow-up was performed if this was considered clinically necessary by the local care team. Lesions showing a  $< 20\%$  increase in size during the 24-month follow-up were considered benign. The participants' clinical notes were reviewed at 24 months to determine how the patient was managed after imaging. This included investigative procedures received, surgical interventions and associated inpatient stays. This was to confirm that decision-tree models accurately reflected clinical practice (e.g. proportion of patients with imaging positive for malignancy undergoing biopsy or surgical excision), and to inform subsequent economic analyses. Follow-up investigations and outcomes for incidental extrathoracic findings on positron emission tomography–computerised tomography and dynamic contrast-enhanced computerised tomography were also collected.

The diagnostic accuracy of the tests was assessed by sensitivity, specificity and accuracy. For the primary outcome, these used prespecified cut-off points, with further exploratory analyses performed considering the full spectrum of cut-off points and the combination of the two tests. Receiver operating characteristic curves were constructed for these exploratory analyses to compare accuracy. For positron emission tomography–computerised tomography, a test was considered positive if it met one of the following criteria: nodule tracer uptake equal to or greater than that of the mediastinum with irregular/spiculate morphology on computerised tomography, or evidence of distant metastases on positron emission tomography or computerised tomography. For dynamic contrast-enhanced computerised tomography, an enhancement threshold of  $> 20$  Hounsfield units was considered malignant.

The outcome measures used in the economic model-based cost-consequences analysis compared positron emission tomography-computerised tomography, dynamic contrast-enhanced computerised tomography, and dynamic contrast-enhanced computerised tomography plus positron emission tomography-computerised tomography in terms of accuracy, estimated life expectancy and quality-adjusted life-years. Costs were estimated from an NHS and Personal Social Services perspective. The incremental cost per malignant case treated and the incremental cost per correctly managed case were also estimated.

### ***Economic evaluation***

The decision-analytic model on which the cost-consequences and cost-effectiveness analyses were based was developed to synthesise evidence and to estimate the expected costs and consequences of each imaging strategy for a cohort of people aged 69 years presenting with a solitary pulmonary nodule (of 8–30 mm) and managed according to the imaging test result. The time horizon of the model was 2 years, but life expectancy and quality-adjusted life-years were extrapolated over the patient lifetime.

Imaging test accuracy and probabilities of following different management pathways were sourced from the trial, the literature and clinical expert opinion. Cost estimates were derived from routine sources (i.e. NHS reference costs, Personal Social Services costs), as well as from the literature, and were inflated, when necessary, to 2018 prices. Further evidence required to estimate life expectancy and health-related quality of life was sourced from the literature.

Parameter uncertainty within the model was addressed using probabilistic sensitivity analysis. Multiple univariate one-way sensitivity analyses were used to identify those parameters to which costs and the proportion of accurately treated cases were most sensitive. Scenario analyses explored the impact of structural assumptions (i.e. exclusion of indeterminate results) on the costs and consequences.

Model validation involved the comparison of results with an independent model, developed to answer the same decision question using different software.

## **Results**

### ***Systematic review of economic evaluation***

Searches identified 664 candidate publications, with nine studies included. These primarily assessed the cost-effectiveness of watchful waiting, computerised tomography, fluorodeoxyglucose-positron emission tomography and surgery, reporting a range of outcomes (i.e. diagnostic accuracy, life-years and quality-adjusted life-years). The studies were judged to be of poor quality when assessed against current methodological guidance for economic evaluations of diagnostic interventions. Although the review provided some evidence of the cost-effectiveness of strategies in relation to the probability of malignancy, it was not possible to make quantitative conclusions because of the heterogeneity among the studies. Given the lack of good-quality evidence on the cost-effectiveness of diagnostic strategies for characterising solitary pulmonary nodules, this systematic review identified the need for a full economic evaluation as part of the trial.

### ***Multicentre comparative accuracy trial***

Of the 380 participants recruited, 312 (47% female, median age of 69 years, age range 35–89 years) completed the study with matched dynamic contrast-enhanced computerised tomography and positron emission tomography-computerised tomography examinations. A total of 57% reported being ex-smokers, and 25% reported being current smokers. The median pulmonary nodule diameter on baseline computerised tomography scans was 15 mm (interquartile range 12–20 mm). There was a 61% rate of malignancy at 2 years. These cancers were found to be 76% non-small-cell carcinoma, of which 74% were adenocarcinomas and 21% were squamous cell carcinoma; 6% carcinoids; and 4% small-cell lung cancer.

The sensitivity and specificity for dynamic contrast-enhanced computerised tomography were 95.3% (95% confidence interval 91.3% to 97.5%) and 29.8% (95% confidence interval 22.3% to 38.4%), respectively; for positron emission tomography-computerised tomography grade, the sensitivity and specificity were 72.8% (95% confidence interval 66.1% to 78.6%) and 81.8% (95% confidence interval 74.0% to 87.7%), respectively. The area under the receiver operating characteristic curve was 0.62 (95% confidence interval 0.58 to 0.67) for dynamic contrast-enhanced computerised tomography and 0.77 (95% confidence interval 0.73 to 0.82) for positron emission tomography-computerised tomography ( $p < 0.001$  for difference). Using a quantitative metric of nodule uptake on positron emission tomography-computerised tomography with a maximum standardised uptake value of  $\geq 2.5$  as a cut-off point was no more accurate than the combined positron emission tomography-computerised tomography grading (area under the receiver operating characteristic 0.79, 95% confidence interval 0.74 to 0.83;  $p = 0.5177$  for difference). Exploratory modelling of the various parameters at different thresholds showed that the maximum standardised uptake value had the best diagnostic accuracy, with an area under the curve of 0.87 (95% confidence interval 0.83 to 0.91), which increased if combined with dynamic contrast-enhanced computerised tomography peak enhancement, with an area under the receiver operating characteristic curve of 0.90 (95% confidence interval 0.86 to 0.93). These exploratory models suggest potential sensitivity and specificity values of 80.5% (95% confidence interval 74.3% to 85.5%) and 78.2% (95% confidence interval 69.9% to 84.6%), respectively, for the maximum standardised uptake value and of 84.7% (95% confidence interval 78.8% to 89.1%) and 77.3% (95% confidence interval 69.0% to 83.9%), respectively, for the combination with peak enhancement.

Dynamic contrast-enhanced computerised tomography was, on average, less costly (£3305, 95% confidence interval £2952 to £3746) than positron emission tomography-computerised tomography (£4013, 95% confidence interval £3673 to £4498) or dynamic contrast-enhanced computerised tomography plus positron emission tomography-computerised tomography (£4058, 95% confidence interval £3702 to £4547). Positron emission tomography-computerised tomography resulted in more correctly managed malignant cases (0.44, 95% confidence interval 0.39 to 0.49) than dynamic contrast-enhanced computerised tomography (0.40, 95% confidence interval 0.35 to 0.45). However, dynamic contrast-enhanced computerised tomography plus positron emission tomography-computerised tomography further improved this proportion to 0.47 (95% confidence interval 0.42 to 0.51). In a sensitivity analysis, the findings of positron emission tomography-computerised tomography being more effective for the correct management of cases and dynamic contrast-enhanced computerised tomography being more cost-effective were robust to varying the model parameters over a range of 50%. In the incremental cost-effectiveness analyses, the cost-effectiveness acceptability curves showed that dynamic contrast-enhanced computerised tomography combined with positron emission tomography-computerised tomography is more likely to be cost-effective at willingness-to-pay thresholds of  $> \text{£}11,395$  per malignant case treated and  $\text{£}11,323$  per accurately managed case. It was not possible to follow up all incidental findings to determine if these were pathologically confirmed or incidental findings. As a result, a health economic analysis of this was not performed.

Symptom analysis showed that unexpected tiredness first experienced in the previous 3 months and more colds or flu in the previous 12 months were positively associated with lung cancer.

## Discussion

In this trial, we have found that positron emission tomography-computerised tomography is the more accurate technique for the diagnosis of solitary pulmonary nodules. However, despite its slightly poorer performance, dynamic contrast-enhanced computerised tomography is a more cost-effective strategy. When the willingness-to-pay threshold per correctly treated malignancy was  $< \text{£}9000$ , dynamic contrast-enhanced computerised tomography was always the preferable strategy. However, when society's willingness to pay for one more correctly treated malignancy increased to  $\text{£}16,000$ , the strategy that combines dynamic contrast-enhanced computerised tomography with positron emission tomography-computerised tomography becomes the strategy most likely to be considered cost-effective, with a probability equal to 1.

These findings have significant implications for the NHS, especially in the light of the introduction of computerised tomography-based lung cancer screening. The availability of positron emission tomography-computerised tomography is limited to relatively few centres and is constrained by the national positron emission tomography contract. The fluorodeoxyglucose radiotracer is produced off-site and supply can be unreliable, resulting in delayed or postponed examinations. In comparison, computerised tomography machines are widely available and relatively inexpensive, and a contrast examination is quick to perform. This raises the possibility of same-day progression from initial nodule detection to subsequent workup with dynamic contrast-enhanced computerised tomography. Case selection could be achieved with artificial intelligence algorithms incorporated into routine screening workstreams.

In this study, the radiation dose to the patient from dynamic contrast-enhanced computerised tomography was higher, at 30 mSv, than a standard positron emission tomography-computerised tomography dose of 14.1 mSv. The protocol was designed 9 years ago and the high radiation dose ensured a minimum signal-to-noise ratio on all scanners, irrespective of manufacturer. However, dual-energy scanning is a routine feature of newer computerised tomography scanners, as is iterative reconstruction, which significantly reduces radiation dose without loss of image quality. A second, and more significant, limitation of dynamic contrast-enhanced computerised tomography in the current study is poor specificity, compared with that expected from the meta-analysis of the current literature. The examinations were not always read by dedicated thoracic radiologists, which is the recommendation for the computerised tomography lung screening programme.

The clinical predictors of unexpected tiredness first experienced in the previous 3 months and more colds or flu in the previous 12 months, which were positively associated with lung cancer, is highly novel, but it is emphasised that these are exploratory findings and of little value until demonstrated in a prospective study of a different data set.

The NHS England lung health check recommendation is that all solitary pulmonary nodules of  $> 5$  mm or  $> 80$  mm<sup>3</sup> on lung settings should be classified indeterminate. Our study used a threshold of 8 mm and was designed well before this lower size threshold was introduced.

## Conclusion

Although positron emission tomography-computerised tomography is the more accurate technique, the low cost of dynamic contrast-enhanced computerised tomography means that dynamic contrast-enhanced computerised tomography alone, or as a gate keeper to positron emission tomography-computerised tomography, is the most cost-effective approach to the diagnosis of solitary pulmonary nodules, albeit with an increased radiation dose. A combination of maximum standardised uptake value and peak enhancement had the highest accuracy, with a small increase in costs. A combined positron emission tomography-computerised tomography and dynamic contrast-enhanced computerised tomography approach with a slightly higher willingness-to-pay threshold to avoid missing small cancers or avoid a watch-and-wait policy should be considered.

The research recommendations are as follows:

- Explore the integration of the dynamic contrast-enhanced component into the positron emission tomography-computerised tomography examination for the characterisation of solitary pulmonary nodules.
- Explore the feasibility of two-stage computerised tomography lung screening with dynamic contrast-enhanced computerised tomography at the same visit if a suspicious nodule is found.

- Undertake analysis of positron emission tomography–computerised tomography and dynamic contrast-enhanced computerised tomography by tumour type, grade and size, using different standardised uptake values and enhancement thresholds to improve accuracy.
- Develop a new protocol for dynamic contrast-enhanced computerised tomography with a lower radiation dose suitable for the newer computerised tomography machines.

## Study registration

This study is registered as PROSPERO CRD42018112215 and CRD42019124299, and the trial is registered as ISRCTN30784948 and ClinicalTrials.gov NCT02013063.

## Funding

This project was funded by the National Institute for Health Research (NIHR) Health Technology Assessment programme and will be published in full in *Health Technology Assessment*; Vol. 26, No. 17. See the NIHR Journals Library website for further project information.



# Chapter 1 Introduction

Parts of this report have been reproduced with permission from Gilbert *et al.*<sup>1</sup> Reproduced from Thorax, Gilbert FJ, Harris S, Miles KA, Weir-McCall JR, Qureshi NR, Campbell Rintoul R, *et al.*, 2021, with permission from BMJ Publishing Group Ltd.

## Background and rationale

Although the incidence of lung cancer is slowly reducing in the UK, the number of new cases diagnosed each year is > 47,000.<sup>2</sup> A proportion of patients with lung cancer present with a solitary pulmonary nodule (SPN) on diagnostic imaging tests; these patients form an important subgroup, as early-stage disease has excellent survival rates following surgical resection.<sup>3</sup> However, not all SPNs are due to lung cancer and the accurate characterisation of SPNs is an ongoing diagnostic challenge with significant associated health costs. A 2010 observational study<sup>4</sup> found that the average US Medicare expenditure for clinical management of a patient with a SPN was US\$50,233 (£30,363) when the SPN was malignant and \$22,461 (£13,577) when it was benign. With the advent of NHS England's Targeted Lung Health Checks programme,<sup>5</sup> piloting low-dose computerised tomography (CT) lung cancer screening in England, the number of patients with a SPN requiring further investigation will increase substantially. A previous Health Technology Assessment (HTA) review<sup>6</sup> noted that CT screening is associated with a relatively high false-positive rate and that subsequent investigations constitute a significant cost. Furthermore, SPNs are a common finding in whole-body screening CT examinations offered to asymptomatic individuals by independent sector providers. Typically, the costs of follow-up investigations from these examinations are incurred in the public sector (UK National Screening Committee, appendix 17). Novel cost-effective approaches to the assessment of SPNs would be of value to the NHS.

## Imaging techniques

The presence of calcification in a SPN on a CT scan is strongly predictive of a benign cause. However, morphological features used to evaluate non-calcified SPNs by conventional CT show considerable overlap between benign and malignant nodules. Widely adopted clinical guidelines<sup>8,9</sup> for the investigation of SPNs recommend serial CT scans for nodules of  $\leq 8$  mm in diameter to look for growth. For nodules of  $> 8$  mm, the recommendation is to perform one of the following: short-term interval CT scans or fluorine-18-labelled-fluorodeoxyglucose (<sup>18</sup>F-FDG) positron emission tomography-computerised tomography (PET/CT) (referred to hereafter as PET/CT rather than <sup>18</sup>F-FDG-PET/CT) and/or biopsy, depending on local expertise.<sup>10</sup>

In the UK, the National Institute for Health and Care Excellence (NICE) currently recommends PET/CT for the assessment of SPNs in cases when a biopsy is not possible or has failed, depending on nodule size, position and CT characterisation.<sup>11</sup> PET/CT acquires images of the body following intravenous injection of a radioactive glucose analogue. PET/CT characterises SPNs on the basis of uptake of glucose. Radionuclide uptake can be assessed qualitatively or quantitatively, with a standardised uptake value (SUV) of  $> 2.5$  implying malignancy. A 2018 meta-analysis<sup>12</sup> confirmed the accuracy of PET/CT as a non-invasive means of characterising SPNs, with a pooled sensitivity of 89% [95% confidence interval (CI) 87% to 91%] and specificity of 70% (95% CI 66% to 73%). In a 2013 audit of a local PET/CT service serving a population of 1 million, with an annual lung cancer incidence of 695 patients, 44 PET/CT scans were requested per year to characterise SPNs.<sup>13</sup> Extrapolated to the UK population, the present demand for PET/CT to characterise SPNs is  $\approx 2700$  examinations per year, equivalent to almost 15% of NHS-funded PET/CT scans performed annually in the UK.

Dynamic contrast-enhanced computerised tomography (DCE-CT) is a rapid series of CT images acquired following intravenous administration of conventional iodinated contrast media. In comparison with PET/CT, which characterises nodules based on their metabolic activity, DCE-CT characterises SPNs on the basis of their vascularity, measured through the amount of enhancement and wash-out.

A 2008 meta-analysis<sup>14</sup> identified 10 studies reporting the ability of DCE-CT to characterise SPNs, with a pooled sensitivity of 93% (95% CI 88% to 97%) and specificity of 76% (95% CI 68% to 97%). Owing to the low cost, high sensitivity and negative predictive value (NPV) of DCE-CT, this technique may be particularly valuable in the assessment of non-calcified SPNs in patients who have a low prior probability of malignancy. Despite the comparable diagnostic accuracy in meta-analyses, however, DCE-CT is not widely used in the UK.

As identified in a 2006 HTA review,<sup>6</sup> there are no agreed guidelines for the further diagnostic investigation of SPNs identified in a CT lung cancer screening programme. The current NICE guidelines<sup>11</sup> for the diagnosis and management of lung cancer were constructed for patients presenting symptomatically or incidentally, and modifications are now required, with CT screening likely to be adopted in the future. The prevalence of malignancy among positive screenings (1.8–3.2%) is significantly lower than for SPNs presenting clinically, for which the rate is closer to 48%.<sup>6,15</sup> As a result, NICE has suggested that imaging approaches may be more appropriate than biopsy for low-risk patients; therefore, imaging approaches are likely to be particularly valuable for the assessment of SPNs identified in a CT screening programme. The results of the UK lung cancer screening randomised controlled trial<sup>16</sup> have demonstrated the potential of a CT screening programme to detect early-stage lung cancer and deliver curative treatment in a high-risk cohort. This HTA-funded pilot study concluded that a single screening in those 60–75 years would be clinically effective and cost-effective at reducing lung cancer-related mortality among high-risk smokers. The incremental cost-effectiveness ratio (ICER) of screen detection, compared with symptomatic detection, was £6325 per life-year gained, and £8500 per quality-adjusted life-year (QALY) gained. Among this cohort, 27% required either early CT follow-up or referral to a multidisciplinary team (MDT) for further workup, and 3.2% had nodules of  $\geq 10$  mm that were suitable for PET/CT or DCE-CT analysis.

The English 14-site pilot of low-dose CT screening is inviting  $> 600,000$  people aged 55–74 years, over 4 years, who have a high risk of lung cancer as result of their smoking status. Based on this, this cohort alone could annually yield  $\approx 4500$  additional non-calcified SPNs, of  $\geq 10$  mm in size, suitable for imaging evaluation with PET/CT or DCE-CT. This demand would represent a significant additional burden on the currently limited availability of PET/CT in the UK, which could potentially be reduced by adoption of management strategies incorporating DCE-CT.

To date, only three studies<sup>17–19</sup> have directly compared the diagnostic performances of PET/CT and DCE-CT in the same cohort of patients. Pooled data from these studies (217 SPNs) indicate that PET/CT and DCE-CT have sensitivities of 92% and 87%, and specificities of 90% and 83%, respectively. As of yet, no comparative studies of PET/CT [neither dedicated positron emission tomography (PET) nor PET/CT] and DCE-CT have been performed in the UK, to our knowledge. Therefore, in this study, objective 1 was to determine, with high precision, the diagnostic performances of DCE-CT and PET/CT in the NHS for the characterisation of indeterminate SPNs.

### ***Therapeutic impact and cost-effectiveness of imaging for solitary pulmonary nodules***

A single study<sup>19</sup> has included an assessment of the therapeutic impact of PET/CT in the characterisation of SPNs. This study found that PET/CT either contributed to or was very important in reaching management decisions in 31 out of 112 cases (28%).<sup>19</sup> There have been several studies evaluating the cost-effectiveness of management strategies that include PET/CT for the characterisation of SPNs, in comparison with conventional CT-based watch-and-wait strategies.<sup>20–25</sup> These studies indicate that PET/CT is either cost-saving or cost-effective across several health-care systems for a wide range of prior probabilities of malignancy. A range of effectiveness measures have been adopted in these studies, including accuracy of management, life expectancy and, in one case, quality-adjusted life expectancy. In general, strategies with PET/CT and those without PET/CT have demonstrated similar effectiveness, but with significant differences in cost. However, these studies used neither diagnostic performance data derived from integrated PET-CT systems nor NHS cost structures.

A single study<sup>26</sup> has compared the cost-effectiveness of strategies that include DCE-CT with that of conventional CT- and PET-based strategies. DCE-CT was found to offer a potentially cost-effective diagnostic approach, with savings of up to £2000 per patient compared with conventional CT-based strategies. Furthermore, a strategy in which patients underwent PET/CT only if the DCE-CT result was positive for malignancy was consistently less expensive, but with similar effectiveness when compared with a PET/CT-based strategy. The cost benefits of DCE-CT were greatest when the prevalence of malignancy was low; therefore, this approach may be particularly advantageous in the evaluation of SPNs found during CT screening. However, the analysis in this study was limited by the lack of direct comparative diagnostic accuracy data for DCE-CT and integrated PET/CT at the time of writing, as well as the omission of final outcome measures. Using the diagnostic performance data obtained in fulfilling objective 1, we undertook decision-analytic modelling to assess the likely costs and health outcomes resulting from the incorporation of DCE-CT into management strategies for patients with SPNs (objective 2).

### ***Incremental value of incorporating computerised tomography appearances of a solitary pulmonary nodule into the interpretation of integrated positron emission tomography-computerised tomography***

Previous economic evaluations of PET/CT for SPNs have been based on diagnostic performance data for dedicated PET systems, rather than integrated PET/CT.<sup>20–22,24–28</sup> Two studies<sup>29,30</sup> have shown a small incremental improvement in the diagnostic performance of PET/CT, compared with PET alone, in the characterisation of SPNs. Incorporating the CT appearances of the nodule into the diagnostic interpretation reduced the false-positive rate for SPNs with moderate fluorodeoxyglucose (FDG) uptake, thereby improving diagnostic specificity from 71% to 77%<sup>29</sup> and from 82% to 89%.<sup>30</sup> This incremental improvement in diagnostic accuracy has the potential to affect the cost-effectiveness of PET/CT, but, at the start of this study, had not been demonstrated in an NHS setting, to our knowledge. Therefore, a secondary objective of this study was to assess, in an NHS setting, the incremental value of incorporating the CT appearances of a SPN into the interpretation of integrated PET/CT examinations.

### ***Combined dynamic contrast-enhanced computerised tomography and positron emission tomography-computerised tomography***

Current integrated PET-CT systems allow for the performance of both PET/CT and DCE-CT in a single examination.<sup>31</sup> None of the currently published studies comparing these techniques in the assessment of SPNs has proposed diagnostic criteria that combine features from both modalities, and discrepant cases are poorly reported.<sup>17,18,32–37</sup> It is feasible that combined parameters of FDG uptake and contrast enhancement could improve the diagnostic performance of PET/CT by discriminating between benign and malignant nodules with mildly increased FDG uptake (i.e. SUV of 2.5–4.9). From the few data currently available, inflammatory nodules with moderate FDG uptake would be likely to exhibit higher FDG uptake/contrast-enhancement ratios than malignant nodules. Furthermore, the NPV of a benign result on both PET/CT and DCE-CT could be sufficiently strong to reduce the need for subsequent imaging surveillance. Thus, a further secondary objective of this study was to assess whether or not combining DCE-CT with PET/CT is more accurate and/or more cost-effective in the characterisation of SPNs than either test used alone or in series.

### ***Incidental extrathoracic imaging findings***

Incidental extrathoracic findings are not uncommon in both the PET and CT components of PET/CT examinations performed for thoracic malignancy.<sup>38</sup> These incidental abnormalities have the potential to add to the health outcomes and cost implications of the use of PET/CT in the characterisation of pulmonary nodules, but would remain undetected by DCE-CT, for which image acquisition is limited to the nodule itself. To date, economic evaluations of PET/CT in the characterisation of SPNs have not included this potential impact.<sup>20–22,24–28</sup> Therefore, an additional secondary objective of this study was to document the nature and incidence of incidental extrathoracic findings on PET/CT scans, undertaken for the characterisation of SPNs, and to model their impact on cost-effectiveness.

### ***Rationale***

Solitary pulmonary nodules form a substantial investigative burden, which is likely to increase in the wake of recent positive lung screening trials. PET/CT, although currently the main investigative strategy for early-stage lung cancer and pulmonary nodules, is a limited and expensive resource. Alternative strategies using current NHS infrastructure, such as DCE-CT, may provide a more streamlined and cost-effective diagnostic strategy. However, a large multicentre trial comparing DCE-CT with PET/CT was required to assess the diagnostic and economic validity of such an approach.<sup>39</sup>

### ***Aims and objectives***

The aims of this study were to compare the diagnostic accuracy of DCE-CT with that of PET/CT for the detection of malignancy in SPNs, and the comparative cost-effectiveness of a diagnostic strategy involving either or both of these imaging techniques.

### **Primary objectives**

- To determine, with high precision, the diagnostic performances of DCE-CT and PET/CT in the NHS for the characterisation of SPNs.
- To use decision-analytic modelling to assess the probable costs and health outcomes resulting from incorporating DCE-CT into management strategies for patients with SPNs.

### **Secondary objectives**

- To assess, in an NHS setting, the incremental value of incorporating the CT appearances of a SPN into the interpretation of integrated PET/CT examinations.
- To assess whether or not combining DCE-CT with PET/CT is more accurate and/or cost-effective in the characterisation of SPNs than either test used alone or in series.
- To document the nature and incidence of incidental extrathoracic findings on PET/CT and DCE-CT undertaken for the characterisation of SPNs, and to model their cost-effectiveness.

## Chapter 2 Observational study methods

### Trial design

The trial was designed in accordance with the guidance for the methods of technology appraisal issued by NICE<sup>40</sup> and adopted by NICE in formulating its guidance for the use of PET in the staging of lung cancer.

The trial was designed as a prospective observational study to assess the diagnostic performance and incremental value of DCE-CT by the addition of this modality to PET/CT in a cohort of 375 patients with a SPN, with the trial protocol previously published.<sup>39</sup> The trial flow diagram is presented in *Figure 1*.

### Participants

#### Eligibility criteria

#### Inclusion criteria

- A soft-tissue solitary dominant pulmonary nodule of  $\geq 8$  mm and  $\leq 30$  mm on the axial plane:
  - measured on lung window using conventional CT
  - no other ancillary evidence strongly indicative of malignancy (e.g. distant metastases or unequivocal local invasion).
- If clinicians and reporting radiologists believe that the patient is being treated as having a single pulmonary nodule and there are other small lesions of  $< 4$  mm that would normally be disregarded, the patient should be included in the trial.
- Nodules already under surveillance could be included, provided that the patient had recently undergone or had scheduled PET/CT.
- Aged  $\geq 18$  years at the time of providing consent.
- Able and willing to consent to the study.

#### Exclusion criteria

- Pregnancy.
- History of malignancy in the previous 2 years.
- Confirmed aetiology of the nodule at the time of the qualifying CT scan. As this was a diagnostic study, should the aetiology of the nodule be confirmed by investigation such as PET/CT or bronchoscopy prior to consent, the patient remained eligible, as the decision to include is made on the analysis of the qualifying CT scan.
- Biopsy of nodule before DCE-CT.
- Contraindication to potential radiotherapy or surgery.
- Contraindication to imaging techniques (assessed by local practice).

All patients meeting the inclusion criteria and none of the exclusion criteria were eligible and were recruited consecutively to the study. In giving consent, they were expected to follow the procedures summarised in *Table 1*.

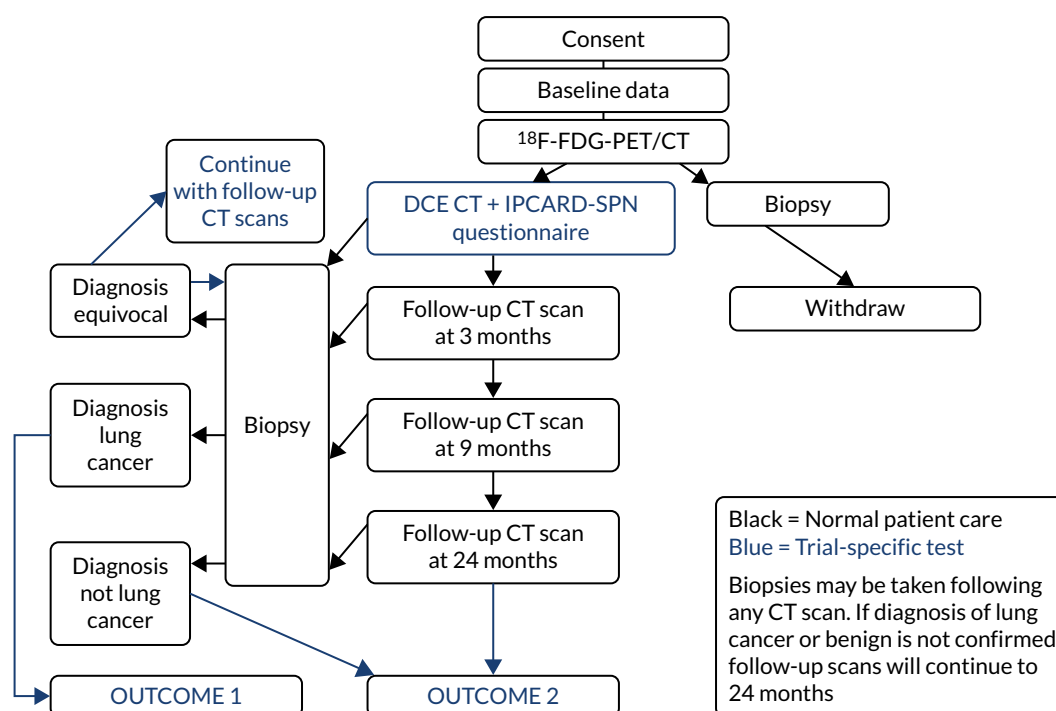


FIGURE 1 Flow diagram of the SPUtNik trial. IPCARD-SPN, Identifying symptoms that Predict Chest And Respiratory Disease – Solitary Pulmonary Nodule; SPUTniK, Single PULmonary Nodule Investigation.

### Changes to eligibility criteria after commencement of the trial

As noted previously, changes to inclusion and exclusion criteria were as follows:

- Inclusion criteria – to include nodules that were already under surveillance, and single nodules when smaller lesions of < 4 mm exist, but they would normally be disregarded by the radiologist.
- Exclusion criteria – ensuring that the criterion for entering the trial is unknown aetiology at the time of CT.

### Setting and recruitment pathway

Patients were identified at local MDT meetings, at the time of referral for investigation of a SPN, or at referral to the PET centres for PET/CT on the basis of having a single dominant pulmonary nodule on a CT scan with uncertain aetiology.

An invitation letter and patient information sheet were sent to potential patients, along with their PET/CT appointment letter, inviting them to participate in the study.

Local research and NHS staff approached potential patients either in clinic or by telephone to:

- explain the study and/or provide the patient information sheet
- note the age, sex and smoking history of a patient
- confirm eligibility for the study.

Patients were given an appointment for PET/CT, and were booked for DCE-CT on either the same day or within 14 days of the PET/CT appointment. (Note that, if there were scheduling issues, appointments could be up to a maximum of 21 days apart.) Some sites chose to make the DCE-CT appointment at the time of the PET/CT appointment, following consent, if there were constraints on scanner time.

TABLE 1 Schedule of observations and procedures

	Screening and recruitment visit(s)	Baseline and diagnostics visit 1 <sup>a</sup>	Visit 2 <sup>b</sup>	Visit 3 <sup>b</sup>	Visit 4 <sup>c</sup>	Visit 5
Day(s)	-14 to -1	0				
Month(s), range			3 months or local practice	9 months or local practice	12–18 months	24 months
Information sheet provided	X					
Informed consent		X				
Review inclusion/exclusion criteria	X	X				
Recruit to study		X				
Check contraindications of contrast	X	X				
4- to 6-hour fasting glucose <sup>d</sup>		X				
Resource assessment						X
Substudy IPCARD-SPN questionnaire		X			X	
PET/CT <sup>d,e,f</sup>		X				
DCE-CT <sup>e,f</sup>		X				
Chest CT <sup>f,g</sup>			X	X		X <sup>c</sup>
Concomitant medications	X					
Adverse events <sup>h</sup>		X				

IPCARD-SPN, Identifying symptoms that Predict Chest And Respiratory Disease – Solitary Pulmonary Nodule.

a Baseline visit consisted of a number of actual visits, depending on if PET and DCE-CT imaging were performed on the same scanner and if consent was taken at the imaging appointment or during a separate appointment.

b Visits 2 and 3 only took place for patients in whom the SPN remained undiagnosed; for patients whose scans were suggestive of lung cancer, visits 2 and 3 may represent biopsy or treatment visits.

c This took place for patients in whom the SPN remained undiagnosed.

d Method in accordance with standard hospital procedure for PET/CT.

e PET/CT and DCE-CT scans were performed on the same day or ideally done within 14 days but no longer than 21 days.

f Copies of scan reports were anonymised and sent to SCTU.

g Chest CT scans were performed only on patients without definitive pathological findings.

h Adverse events were recorded from DCE scan date for up to 30 days after DCE-CT scan. No effects of contrast medium were expected after 7 days.

The Single Pulmonary Nodule Investigation (SPUtNik) study patient information sheets and Identifying symptoms that Predict Chest And Respiratory Disease – Solitary Pulmonary Nodule (IPCARD-SPN) questionnaire were given to patients either in clinic or by post.

## Outcomes

### Primary outcome

This study had two co-primary outcomes. The first was to assess the diagnostic test characteristics [sensitivity, specificity, NPV, positive predictive value (PPV) and overall diagnostic accuracy (ODA)] for PET/CT and DCE-CT, in relation to a subsequent diagnosis of lung cancer within a 2-year time frame. The second was to assess the cost-effectiveness of each imaging technique. The outcome measures used in the economic model include accuracy, estimated life expectancy and QALYs. Costs were estimated from an NHS perspective. ICERs (reported as the incremental cost per correctly treated malignancy and the incremental cost per correctly managed case) compare management strategies with DCE-CT against management strategies without DCE-CT.

### **Secondary outcomes**

Secondary outcome measures include diagnostic test characteristics for PET/CT with incorporation of CT appearances and combined DCE-CT plus PET/CT. The incidence of incidental extrathoracic findings on PET/CT, subsequent investigations and costs were determined.

### **Adverse events**

All adverse events occurring within the 30 days following the study DCE-CT were reported to the SPUtNik database. For full details, see the SPUtNik study protocol.<sup>39</sup>

### **Registration**

Sites registered patients to the SPUtNik study by sending a signed registration form via fax or e-mail attachment to the Southampton Clinical Trials Unit (SCTU). Sites were given a block of SPUtNik codes. Patients were allocated trial identifier codes consecutively, following consent. On registration, SCTU staff checked eligibility and confirmed the patient trial identifier code by e-mail. Registration and DCE-CT could not take place before informed consent was signed; however, registration with SCTU could take place before or after DCE-CT, because there was a possibility that some clinics would schedule the DCE-CT outside office hours.

### **Blinding**

The study was a non-randomised diagnostic accuracy and comparative health economic effectiveness trial; therefore, blinding was not necessary to meet the trial objectives. However, the results of the DCE-CT were not reported to the participants' clinicians, so as not to bias the assessment of the diagnostic accuracy of DCE-CT.

### **Data collection**

#### ***Baseline and evaluative procedures (visit 1, day 0)***

If both imaging techniques were performed on the same day, PET/CT was performed first, with no waiting time between imaging. If they were performed on separate days, either imaging technique could be performed first, provided that patient consent and registration took place before DCE-CT (e.g. in the case of a delay in PET/CT).

Ideally, PET/CT and DCE-CT were performed within 14 days, with an absolute maximum of 21 days allowed between imaging techniques when sites had difficulty with scheduling.

#### ***Follow-up***

Following the PET/CT and DCE-CT investigations, management of the SPN was directed by the local/specialist lung MDT.

In many cases, a nodule biopsy or excision biopsy was undertaken and the histopathological outcome recorded. Cases shown to be due to lung cancer (or other malignancy) were managed according to local protocols. Follow-up/outcome data were collected by case report form (CRF).

Patients with a high pre-test probability of cancer, who were unfit for surgery or for whom a biopsy was non-diagnostic/not possible, were considered for stereotactic ablative radiotherapy (SABR) or nodule ablation. For the purposes of this study, these patients were considered to have cancer as per British Thoracic Society (BTS) guidelines.<sup>8</sup>

In some cases, nodule surveillance was appropriate (with or without prior biopsy).<sup>8,9</sup> In these cases, recommended follow-up was performed at:

- 3 months
- 9 months
- 24 months.

Deviation from these time points, based on clinical need, was at the discretion of the MDT. For instance, if a nodule resolved during follow-up, then continued imaging was not necessary. During the course of the trial, the BTS guidelines changed, such that nodules stable at 12 months on volumetric analysis (< 25% change in volume) are considered benign, requiring no further follow-up. When the technology to undertake this analysis was available to the MDT, follow-up could be terminated before 24 months.

At each study visit, the following were performed:

- chest CT (low dose, thin section, unenhanced), unless the MDT felt that it was inappropriate
- recording of any biopsy samples taken
- recording of health resource use information.

At 2 years:

- Health resource data were collected from patient records (this included additional findings that came to light while investigating the SPN with PET-CT, treatment associated with the SPN and tests, and treatment related to these additional intra- or extrathoracic findings).
- The end-of-study CRF was completed in the clinical database.
- The principal investigator (PI) signed off a patient's electronic CRF record and the database was locked for that patient.

### ***Procedure when nodule was reduced in size or not visible on the dynamic contrast-enhanced computerised tomography scan***

Occasionally, a nodule that was eligible (i.e. was 8–30 mm in size) on the lung window on the qualifying CT scan had reduced in size or was not visible when DCE-CT was performed. This was likely to be related to the inflammatory/infective nature of the nodule. In this case, the following procedures were undertaken:

- If the nodule was not visible on the DCE-CT locating scan, contrast was not administered. If the nodule had reduced in size but was still > 8 mm, contrast was given.
- If the nodule was < 8 mm, a local decision was made by the supervising radiologist.

### ***Withdrawal criteria***

A patient could withdraw consent at any time and was not asked to give a reason:

- If a patient withdrew from undergoing PET-CT or DCE-CT or both, but did not specifically withdraw consent to collect data from hospital notes, collection of relevant data from their hospital notes and general practitioner contact continued.
- If consent was completely withdrawn, results were recorded on CRFs for procedures performed prior to the withdrawal of consent.

## **Sample size**

For the primary outcome measures in this study, the diagnostic characteristics for DCE-CT and PET/CT were used for the sample size calculations. Use of the other outcome measures that are related to the economic analyses was prevented by the prior need for detailed characterisation of the decision trees.

We consider the sample size needed to detect particular accuracy for each test, separately, and then when the tests are used in conjunction.

### **Consideration for each test separately**

Published sensitivity for PET/CT for the characterisation of SPNs varies between 77% and 96% (pooled weighted average: 92%) and specificity varies between 76% and 100% (pooled weighted average: 90%).<sup>14,17,18,21,23,29,30,32,41,42</sup> Published sensitivity and specificity values for DCE-CT vary between 81% and 100% (pooled weighted average: 87%) and between 29% and 100% (pooled weighted average: 83%), respectively.<sup>14,17,18,21,23,29,30,32,41,42</sup> Based on two previous UK studies, the mean prevalence of malignancy in indeterminate SPNs has been reported as 68.5%.<sup>22</sup> At this prevalence, a sample size of 375 participants will produce approximately 257 malignant and 118 benign SPNs. This gives 95% confidence limits for the sensitivity and specificity of DCE-CT of  $87\% \pm 4.1\%$  and  $83\% \pm 6.8\%$ , respectively, with sensitivity and specificity values for PET/CT of  $92\% \pm 3.3\%$  and  $90\% \pm 5.4\%$ , respectively. These estimates will provide sufficiently narrow confidence limits to allow precise economic modelling based on the results. For the purposes of economic analyses, we also considered combining our data with the meta-analysis results from our systematic review (see *Chapter 8*) of previous studies of 217 patients who had undergone both techniques.<sup>12,14,15</sup> Recruitment rates were anticipated to be high (70%) because DCE-CT was additional, rather than an alternative, to normal care, and is readily incorporated into the existing PET-CT examination. We expected to recruit the required sample size ( $n = 375$ ) in 18 months. Assuming that only 70% of patients would meet all inclusion criteria, it was anticipated that we would need to screen  $375/0.7 = 536$  patients.

### **Consideration for when both tests are used together**

Consideration has to be given to both tests being used together. In particular, (1) those with a negative result on a DCE-CT scan are classed as benign, and (2) those with a positive result on a DCE-CT scan then undergo PET/CT; those with a positive result on the PET/CT are classed as 'malignant' and those with a negative result on a PET/CT scan are classed as 'benign'. The specificity of this process is the same as that of using PET/CT alone. So, the key interest is estimating the sensitivity of this joint-test classification strategy, compared with PET/CT alone. Based on previous data of 130 malignant tumours, 114 tumours give a positive DCE-CT scan result and a positive PET/CT scan result; this suggests that the sensitivity of the joint testing procedure is  $114/130 = 0.877$ . Compared with the PET/CT sensitivity, thought to be 0.92 (as noted previously), the joint-testing approach is projected to reduce sensitivity by about 4%. Based on the sample size formula of Alonzo *et al.*,<sup>43</sup> to detect that the joint DCE-CT-PET/CT approach has at least a 4% reduction in sensitivity, compared with the PET/CT sensitivity of 0.92, a total sample size of 288 participants is required (including 197 with a malignant tumour); this calculation assumes an 80% power, 5% significance level, and prevalence of malignancy of 0.685. Thus, by including 375 participants, as per our previous sample size calculations, our study is also powered to detect at least a 4% decrease in sensitivity for the joint-testing approach.

## **Statistical methods and data analysis**

We considered the diagnostic accuracy of positive PET/CT and DCE-CT scan results, both separately and in conjunction, in relation to a diagnosis of lung cancer by 2 years. In these analyses, we will be able to include only those patients in whom the presence of lung cancer by 2 years is confirmed.

Initially, the separate diagnostic performance of PET/CT and DCE-CT were examined using the predefined diagnostic criteria. For PET/CT, this was based on the combined PET grade and CT grade (see *Table 3* for grade breakdown). This was classified as positive for malignancy if one of the following criteria was met: grade 4 on PET or CT, or grade 3 on both PET and CT, or grade 2 on PET and grade 3 or 4 on CT. For DCE-CT, 15-Hounsfield unit (HU) enhancement is prevalent in the literature when scanning patients at an X-ray energy of 120 kVp. We proposed scanning patients at 100 kVp to obtain a higher signal-to-noise ratio. The attenuation of iodine increases with decreasing photon energy. Phantom measurements verified that an iodine concentration measuring 15 HU at 120 kVp measured approximately 20 HU at 100 kVp; as a result, 20 HU was chosen as the prespecified threshold for malignancy.

In addition, we examined the full range of possible threshold values to see if performance could be improved. This involved using the grading of individual CT and PET images, as well as the maximum and mean SUV from the PET image, peak enhancement, peak and mean HU from the DCE-CT and the reporting radiologist's classification of the SPN from the PET/CT and DCE-CT results. The classification produced by the site radiologists was their expert opinion and was based on all available information from either the PET/CT or DCE-CT (correspondingly), as well as any prior information from the initial staging CT. An indeterminate option was available here. At each threshold, sensitivity and specificity were estimated (with 95% CIs), and the receiver operating characteristic (ROC) curve calculated. The optimal cut-off point from the range of values is reported based on keeping the sensitivity above 90% and maximising specificity within this limitation. An alternative cut-off point that provides the best trade-off in sensitivity and specificity is also reported. The sensitivity, specificity, PPV, NPV and ODA at these cut-off points, as well as the pre-defined values, are presented with 95% CIs in this report. When translating sensitivity and specificity to PPVs and NPVs, we assumed a particular prevalence, using the value seen in our study. We report PPVs and NPVs for a range of other prevalences reported in the literature, identified from our systematic review. The ODA is the percentage of cases that are correctly classified, regardless of whether they were cancers or non-cancers.

The diagnostic performance of PET/CT and DCE-CT combined was examined using the same techniques as previously described, with patients classed as 'positive' if they had both PET/CT and DCE-CT positive scan results, and all other patients classed as 'negative'.

A logistic regression model was undertaken, including key diagnostic elements from both PET/CT and DCE-CT. This included (but was not limited to) the individual imaging components that were described previously, size measures from each of the scans and predefined measures such as the standardised perfusion value (SPV). The SPV is the maximum enhancement multiplied by the subject's body weight and divided by the dose of iodine received. These variables were used on their original scale as covariates in a logistic regression to produce a risk score and predicted probability of lung cancer for each individual, based on their specific test values. This predicted probability came from a transformation of the linear predictor from the logistic regression model: the predicted 'risk' for each individual. A cut-off value was used to decide a high-risk score (which predicts an adverse outcome) and a low-risk score (predicting a good outcome). The calibration of the model was assessed by grouping patients into deciles ordered by predicted risk and considering the agreement between the mean predicted risk and the observed number of true lung cancer cases in each decile (sometimes referred to as the expected vs. observed statistic). The derived diagnostic rule was cross-validated by comparing the classification of each patient with their outcome of confirmed lung cancer, allowing an estimate of the sensitivity and specificity of the prediction model. By varying the chosen cut-off level, a ROC curve was produced, summarising the sensitivity and specificity of the predictive rule across the full range of cut-off points. The overall discriminatory ability was summarised as the area under the receiver operating characteristic (AUROC) curve with 95% CI. The most suitable cut-off level was selected using the same rules as before. The internal validity of the final model was assessed by the bootstrap resampling technique to adjust for overoptimism in the estimation of model performance due to validation in the same data set that was used to develop the model itself. These models will be exploratory only and, before they can be used or considered useful, they would require external validation in a separate data set.

If these methods showed poor accuracy performance (in terms of calibration and/or discrimination), then the logistic regression model was extended to include additional patient-level covariates (such as time from  $^{18}\text{F}$ -FDG injection to PET/CT scan) in addition to test results. Demographic information was considered, as well as clinical and imaging features considered indicative of a higher likelihood of SPN malignancy. The performance of the model was also evaluated at the site level, when possible, to ascertain whether or not model performance is consistent in each site or, if it is not, the variability in performance across sites, and whether or not this can be improved by tailoring the prevalence in each site.

## Interim analysis and data storage

No interim analysis was planned for this study, although there were a series of Data Monitoring and Ethics Committee meetings that occurred throughout the duration of the data collection process.

Data and all appropriate documentation will be stored for a minimum of 15 years after the completion of the trial, including the follow-up period.

## Accreditation of positron emission tomography centres

### Methods

Centres performing PET for the SPUtNik study underwent an accreditation process using procedures established for multicentre trials by the UK PET Core Lab.<sup>44</sup> For centres to be accredited by the UK PET Core Lab, they had to submit the following data:

- PET/CT acquisition and reconstruction of the International Electrotechnical Commission image quality phantom following a standard protocol; when possible, the local clinical protocol was accredited
- submission of 8 weeks' phantom data to demonstrate ongoing scanner stability
- evidence of ongoing PET and CT quality control (QC) as part of a quality assurance (QA) system
- evidence of calibration of the patient weighing scales used to determine SUVs
- evidence of traceability of the radionuclide calibrator used to measure injected activity
- two anonymised patient scans for visual assessment of image quality by two experienced PET clinicians.

To ensure that image quality was maintained throughout the trial, each scanning centre was required to submit a scan of the image quality phantom to the UK PET Core Lab for analysis on an annual basis. In total, 16 hospitals recruited patients as part of the study. Not all recruiting hospitals had access to a fixed PET/CT scanner on-site, so patients were either scanned in a mobile PET/CT unit or sent to the nearest PET centre. *Appendix 2, Table 27*, shows a summary of the recruiting sites and the fixed PET centres ( $n = 17$ ) or mobile units ( $n = 8$ ) used.

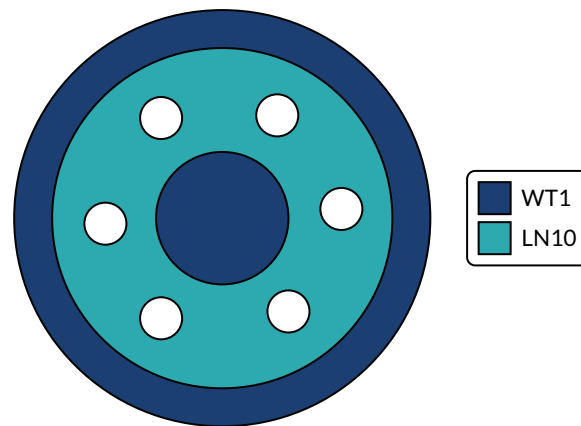
To improve recruitment, patients with nodules already under surveillance were included in the trial if PET had been performed recently. In some cases, this meant that non-accredited PET scanners were used at the time of the PET. In all but one case, the scanners were retrospectively accredited.

## Accreditation of dynamic contrast-enhanced computerised tomography centres

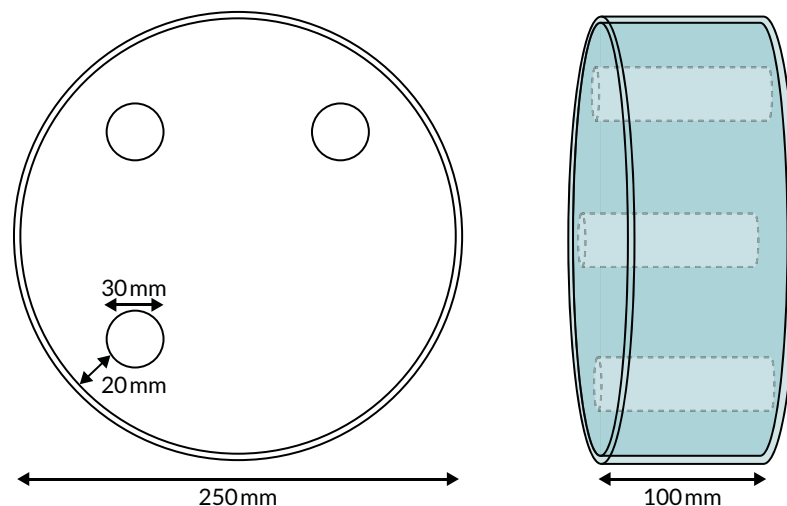
Centres performing DCE-CT for the SPUtNik study underwent an accreditation process developed specifically for this trial and conducted by the Radiation Protection Department of the Mount Vernon Cancer Centre, East and North Hertfordshire NHS Trust.

Two phantom types were designed for the trial. The SPUtNik chest-equivalent phantom (*Figure 2*) was constructed by the Clinical Physics Department at St Bartholomew's Hospital, and the SPUtNik water-filled Radiographer QA Phantom (*Figure 3*) was constructed by the Bioengineering Department at the Mount Vernon Cancer Centre.

A clinical scientist from the DCE-CT accreditation team visited each site and, with the lead radiographer, entered the weight-dependent trial protocols into the scanner and saved them with clearly identifiable names. The chest-equivalent phantom was scanned under the trial protocol (see the SPUtNik study protocol<sup>39</sup>). One SPUtNik water-filled radiographer QA phantom was issued to the centre and scanned,



**FIGURE 2** The SPUtNik chest-equivalent phantom. Outer annulus, diameter 290 mm, composed of water-equivalent material WT1 (Clinical Physics Department, St Bartholomew's Hospital); middle annulus, diameter 240 mm, composed of lung-equivalent material LN10; central circle, diameter 100 mm, composed of water-equivalent material WT1. Six holes, each 30 mm in diameter, designed to hold a vial containing iodine solution at the following concentrations: 0.0, 0.3, 0.6, 1.6, 2.4 and 4.7 mg/ml. Phantom thickness: 100 mm.



**FIGURE 3** The SPUtNik water-filled radiographer quality assurance phantom. Perspex cylinder: 250 mm outer diameter, 244 mm internal diameter. Water filled with release plug to fill/empty phantom. Three internal chambers, each of 30 mm internal diameter, with plugs to fill/empty with iodine solution at the following concentrations: 0.6, 1.6 and 2.4 mg/ml.

as per the QA protocol (see the SPUtNik study protocol<sup>39</sup>). Digital Imaging and Communications in Medicine- (DICOM)-standard images of both phantoms were saved to a compact disc for later analysis. The peak energy (in kVp) and half-value layer of the X-ray beam at 100 kV were measured using the RaySafe™ Xi R/F Detector (RaySafe, Billdal, Sweden), positioned in the scan plane and using a scout projection, taking care to avoid attenuation by the table/couch.

Information and instruction were given to the lead radiographer for QA for the SPUtNik trial and for data anonymisation and transfer to the PET Core Lab for storage and further analysis of patient images. Images exported from each scanner were checked to ensure that the correct scan protocol had been set up. The mean and standard deviation in CT number in the iodine inserts of both the chest-equivalent phantom and the Radiographer QA phantom were extracted using IQWorks [<http://iqworks.org/> (accessed 12 October 2018)]. The line of best fit between CT number and iodine concentration in each phantom was assessed using Microsoft Excel® 2010 (Microsoft Corporation, Redmond, WA, USA), and the gradient was taken as the iodine calibration factor at 100 kV. Sites were issued a certificate of accreditation and informed of the iodine calibration factor in the SPUtNik chest phantom for their scanner.

All scanners were required to undergo regular planned preventative maintenance and CT number calibration, at frequencies recommended by the manufacturer.

Fourteen sites recruited patients as part of the study. Participants underwent DCE-CT on one of 16 scanners, as summarised in *Appendix 2, Table 28*.

The mean iodine calibration factor in the lung-equivalent phantom was 32.2 HU/(mg/ml), with a coefficient of variation across the scanners of 6.6%. The mean iodine calibration factor in the water-filled radiographer phantom was 30.0 HU/(mg/ml), with a coefficient of variation of 4.5%.

### **Dynamic contrast-enhanced computerised tomography acquisition design**

The DCE-CT acquisition settings were chosen to reasonably standardise dose and image quality across all manufacturers and models of scanner. The full procedure is listed in the SPUtNik study protocol.<sup>39</sup>

Computerised tomography scans taken at 120 kVp are prevalent in the literature; however, by scanning at a lower photon energy, the increased attenuation of iodine could be exploited. For a given iodine concentration, the increased attenuation at low energy results in a higher CT number. Scans of phantoms simulating patients of different weight explored the use of 100 kVp and 80 kVp. At 80 kVp, the X-ray tube current required to maintain a desired signal-to-noise ratio was sufficiently high to exclude most CT scanners. However, at 100 kVp, a good signal-to-noise ratio could be achieved for all phantom sizes. Therefore, 100 kVp was chosen.

Tube current modulation alters the X-ray current, and, consequently, the photon intensity, with variation in patient attenuation during CT. The variation in manufacturer approach to tube current modulation would have resulted in differences in dose and image quality, depending on patient weight and scanner type. Standardisation of tube current modulation between manufacturers was beyond the scope of this trial; therefore, fixed tube currents were used across three weight categories: < 60 kg = 200 mA, 60–90 kg = 350 mA and > 90 kg = 500 mA. These tube currents were chosen to maintain a signal-to-noise ratio independent of patient weight.

At the time of trial inception, 64 × 0.6 mm collimation was representative of modern scanners. Slice thickness and interval were chosen to provide adequate noise statistics for the measurement of CT number enhancement, with enough sensitivity to choose the slice position and with maximum enhancement for nodules of > 8 mm in diameter. A reconstruction kernel with similar appearance and noise level across each manufacturer was chosen.

Although some scanners included in the trial had an option of iterative reconstruction, this was not used, so as to standardise the dose and image quality of all CT systems, regardless of scanner model and manufacturer.

## Chapter 3 Site quality control

### Accreditation of positron emission tomography centres and technical quality control of positron emission tomography scans

Imaging protocols with guidance for performing the PET/CT (see SPUtNIK protocol<sup>39</sup>) were provided to sites, with local clinical protocol followed if the PET/CT was carried out as standard of care prior to study entry. Of the 380 patients recruited to the study, 373 were eligible for the study and underwent PET/CT imaging from January 2013 to December 2016. The scans were sent to the UK PET Core Lab, St Thomas' Hospital, for technical checks (Table 2).

Any deviations from the expected patient preparation, image acquisition or scan processing were flagged for discussion with a PET expert. A total of 370 (99.2%) PET scans were submitted for technical review. The results of technical review were as follows:

- Sixteen scans used a new reconstruction not accredited for the study called point-spread function (PSF) modelling.
- Patient blood glucose was not measured/not provided for seven patients; one patient had a blood glucose level of > 11 mmol/l.
- Fasting status was not provided for 27 scans, nine patients fasted for < 6 hours, but all fasted for at least 4 hours and their blood glucose was within the expected range.
- All scans covered at least the angle of jaw to mid-thigh.

The mean injected activity was 352.6 MBq (range 148.1–464.0 MBq). The national diagnostic reference level (DRL) for <sup>18</sup>F-FDG is 400 MBq and it is recommended to use a weight-based protocol of 4.5 MBq/kg, with a minimum injected activity that is dependent on the scanner model and acquisition parameters used. The mean injected activity used for the SPUtNIK patients was 4.9 MBq/kg (range 2.5–12.2 MBq/kg) and patient weight ranged from 30 to 150 kg (mean 76.8 kg). A plot of the injected activity as a function of patient weight is shown in Figure 4.

The mean uptake time was 68.6 minutes (range 49.0 to 117 minutes). The distribution of uptake times is shown in Figure 5; for 72% ( $n = 267$ ) of PET scans, the uptake time was within  $60 \pm 10$  minutes. For 80% ( $n = 295$ ) of scans, the uptake time was within 50–75 minutes. Only one scan had an uptake time of < 50 minutes, and 74 had an uptake time of > 75 minutes. The standard-of-care protocol for one centre was a 90-minute uptake time, rather than the 60-minute uptake time in the other sites.

TABLE 2 Technical checks performed on the PET data

Technical checks performed	Ideal range
Scanner acquisition and reconstruction matched to accredited parameters	Any deviations noted and reviewed by PET expert
Patient blood glucose	< 11 mmol/l
Fasting status	≥ 6 hours
Uptake time	60 minutes ( $\pm 10$ minutes)
Scan range	Angle of jaw to mid-thigh
DCE-CT intravenous contrast	To be administered after PET if performed in the same session
Injected activity	Dependent on scanner and total scan time (diagnostic reference level 400 MBq)

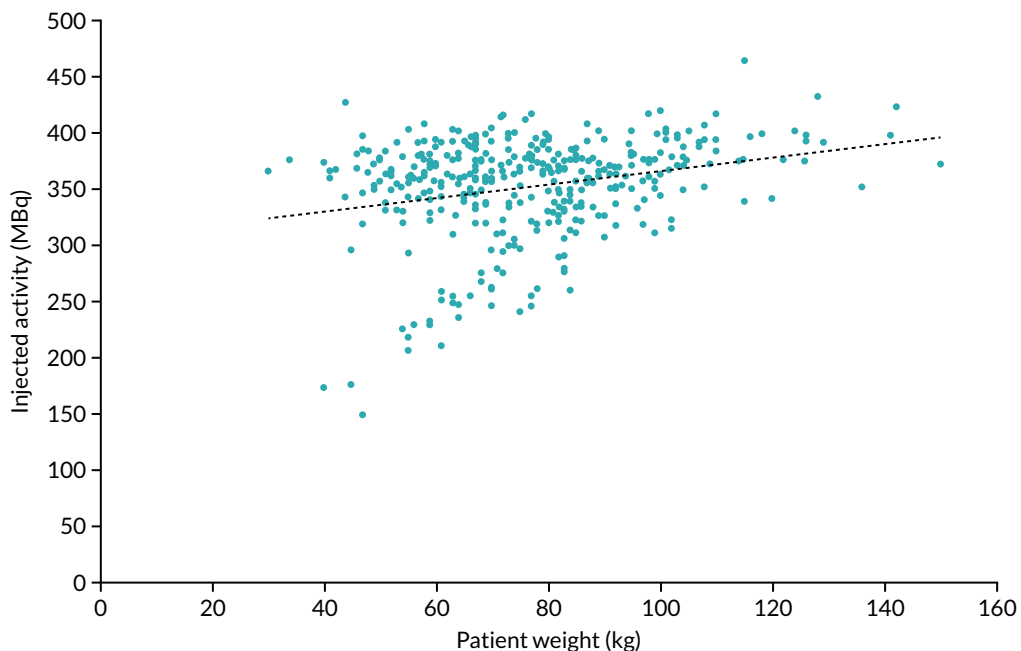


FIGURE 4 Injected activity as a function of patient's weight for all patients.

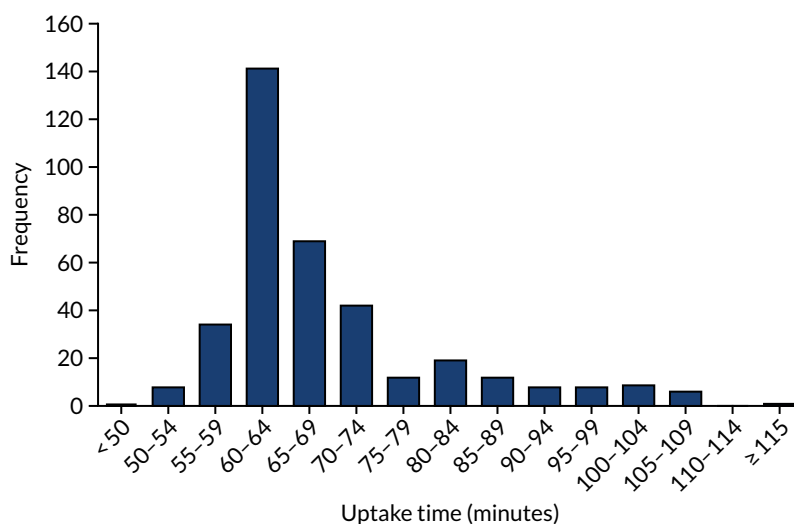


FIGURE 5 Uptake time (time between injection and PET scan start) for all patients.

### Radiation dose

In general, a dose of 7.6 mSv is quoted for the radiotracer injection based on 400 MBq and 6.5 mSv for the CT part, based on national DRLs for CT as part of a PET-CT. This gives a total of 14.1 mSv for the PET-CT. However, local practice varies and the expectation is that the PET injected dose is lower for SPUtNIK, as many sites used weight-based injected activities.

### Radiographer-led quality control of dynamic contrast-enhanced computerised tomography scanners

Each site was issued with a water-filled radiographer QA phantom and shown how to carry out tests on its scanner. Baseline measurements were performed by the clinical scientist, who then created a radiographer QA spreadsheet with tolerances and pass/fail criteria for the site. All results were logged

on the QA spreadsheet, a copy of which was sent to Mount Vernon for analysis at regular intervals throughout the trial. To ensure that QA was not excessively burdensome at participating sites, the frequency of DCE-CT radiographer QA phantom scans was set to either weekly (for sites with a high trial patient volume) or before each patient scan (for less frequent trial scans). QA was not carried out within 7 days of trial DCE-CT on 24 occasions.

Across the 14 sites, QA was carried out a total of 753 times. This ranged from 148 instances at one site to just four at another. Some sites carried out radiographer QA more often than the trial required, whereas other sites, with particularly low recruitment, carried out QA a few times only. In total, there were just six (0.8%) instances when the measured iodine calibration factor fell outside  $\pm 5\%$  of the baseline calibration factor measured at the accreditation visit.

There was no marked variation in iodine calibration factor over time for any of the scanners. The maximum difference between baseline calibration factor (measured at accreditation) and mean calibration factor (over the length of the trial) was  $-2.1\%$ , with the mean variation being  $0.2\%$ . There were no systematic trends.

Sites were asked to record any new tubes that were installed in the scanners. Only one new tube was reported during the trial. The baselines and tolerances for QA measurements were reset following installation of the new tube. Although the measured CT numbers in the phantom did change slightly, the enhancement relative to water was not significantly different and the iodine calibration factor did not change. The cause of the tube change was not ascertained; however, the changes in tube output or beam quality that might be expected prior to the failure of an X-ray tube had no effect on the radiographer QA results or iodine calibration factor. There was one other step-change in CT numbers during the trial. No explanation for this could be found. The enhancement relative to water and iodine calibration factor did not change, so the scanner remained in the trial.

## Technical quality control of dynamic contrast-enhanced computerised tomography scans

### Deviation log

Of the 329 completed DCE-CT sessions carried out during the trial, a number of scans deviated from the scan protocol (see *Appendix 4, Table 30*). Protocol deviations were reviewed by radiologists and clinical scientists to determine if patients should be excluded from the final analysis. The following sections are a list of deviations and the rationale for determining inclusion/exclusion in the trial. The aim was to be pragmatic and inclusive, and to recognise that the imaging was being conducted at multiple NHS sites, reflecting real-life practice. In total, 12 scans were deemed unusable.

### Reconstructed field of view

A significantly larger reconstructed field of view will affect how accurately the region of interest (ROI) is drawn within the nodule and the number of voxels included in the CT number measurement. Four scans had a field of view within a few mm of the 150 mm required in the protocol. This had little effect on the ROI or CT number measurement. These patients were included. Nine scans were reconstructed with a significantly larger field of view (303–426 mm) than specified in the protocol. The nodule diameter for these scans was reviewed against the reconstructed field of view to determine the number of voxels contained within the ROI. If the value was greater than would be measured for an 8-mm nodule with the protocolled 150-mm field of view, the scan data were included for analysis. Five scans were excluded on this basis, as relatively small nodules were scanned on a large field of view. This combination of small nodule and large field of view resulted in fewer voxels sampled for measurement of average CT number, thereby reducing the confidence and accuracy of the CT number measured.

## Reconstructed slice thickness and slice interval

One scan was reconstructed with a slice interval of 2.5 mm, one scan with a slice thickness of 2 mm and 18 scans with an interval of 3 mm (protocol 2.5- or 3-mm slices with a 2-mm interval). This could have a small effect on the choice of central slice of the nodule for CT number measurement. The deviations are not significant, and all scans were included for final analysis.

## Contrast-enhanced imaging times

Contrast-enhanced CT scans should have been acquired at 60, 120, 180 and 240 seconds post iodine injection. Three patients were scanned at incorrect time points. One patient was scanned with an additional 23-second delay to the 2-, 3- and 4-minute time points. Iodine uptake usually takes place within the first minute; the short delay to the final three acquisitions was thought to have little effect on the measured CT number, so this scan was included. A second patient had the 60-second post-contrast scan carried out correctly, but the remaining three acquisitions were carried out at 174, 209 and 270 seconds. The patient effectively had no 2-minute scan, and the peak CT number may have been missed. This scan was excluded from analysis. A third patient had the 60-second post-contrast scan in the wrong location, not including the nodule. The remaining three acquisitions were carried out at the correct times and location. The peak enhancement value was 29.9 HU, which is greater than the predicted 20 HU threshold for malignancy. This scan was included in the analysis.

## X-ray tube current and patient mass

The DCE-CT protocol has three weight-based mA settings: < 60 kg = 200 mA, 60–90 kg = 350 mA and > 90 kg = 500 mA. Seven patients were scanned with the wrong mA value. When a patient received a higher current than intended, the image quality was improved and the scans were included in the trial. When a patient's weight was slightly over or under a weight boundary, the data were included for analysis. One patient was scanned with automated tube current modulation. The mA value delivered for the central slice of the nodule was compared with the protocol mA value. The patient was included, as the enhancement value was significantly lower than the expected threshold for malignancy; therefore, a lower-dose (higher-noise) scan would not have altered the status of this patient's data.

## Radiation dose

A high-dose DCE-CT strategy was used to keep the image noise low and get more accurate enhancement values. Imaging was undertaken at multiple post-contrast time points. The total effective dose for an average patient for the whole DCE-CT examination (whole-chest scan to locate nodule, then one pre-contrast and four post-contrast nodule scans) is  $\approx 30$  mSv. This was calculated using the ImPACT CT dose calculator [[www.impactscan.org/](http://www.impactscan.org/) (accessed 1 October 2018)].

## Reconstruction kernel

Royal Papworth Hospital installed a new Siemens Force CT scanner (Siemens Healthineers AG, Erlangen, Germany) during the trial. The Force CT scanner does not have the same reconstruction kernels as other Siemens scanners used in the trial. During the protocol set-up visit, a variety of reconstruction kernels were reviewed. The kernel with noise statistics most similar to the B30 kernel in the trial protocol was selected.

## Dynamic contrast-enhanced computerised tomography injection rate

Four patients had DCE-CT examinations using a contrast injection rate of 3 ml/second, rather than the 2 ml/second prescribed in the protocol. A linear-systems approach (see *Report Supplementary Material 1*) was used to determine patient-specific correction factors to compensate for this difference in injection rate. The model showed that the faster contrast injection rate could have significantly affected the enhancement values for images acquired at 60 seconds, but the impact on later time points would be minimal. As three of the four cases showed significant nodular enhancement at later time points, it was considered unlikely that their DCE classification as benign or malignant would have been affected by the incorrect contrast injection rate. The remaining cases were excluded from the trial as a result of analysis failure for other reasons.

## Window centre and width

In addition to the deviations above, 19 scans were viewed with a window width of 400 HU (protocol is 350 HU). This has no significant impact on how the nodules are visualised; therefore, all scans were included in the final analysis.

## Measurement software for contrast enhancement

Site accreditation involved measurement of iodine CT number on CT scanner workstations to determine the iodine calibration of each scanner, and also on reporting workstations or picture archiving and communication systems. The measurement of iodine CT number was tested on both CT workstation and reporting workstation to ensure that no additional processing was applied to the images that might affect the trial.

Phantom scans were compared at eight different sites with three different manufacturers of scanner (seven different scanner models) and four different reporting workstation manufacturers. The radiographer QC phantom contains three inserts of varying iodine contrast concentration, equivalent to approximately 20, 50 and 70 HU (at 100 kV) embedded in a larger volume of water. The CT number, standard deviation and enhancement relative to water were compared for measurements taken on the CT scanner and reporting workstation. Measurements for three slices in the phantom were taken, with care taken to avoid air bubbles.

The maximum deviation in CT number for a given slice was 1.4 HU for the highest iodine concentration. At the clinically relevant threshold of 20 HU, the maximum deviation seen was 1.0 HU. There was no overall positive or negative trend in the values, and the differences seen were within the standard deviation (typically 6–10 HU) of the sample of CT numbers within each ROI.

There was no evidence of a difference between the measured CT number and the enhancement value for measurements taken on the CT scanner or reporting workstations.

## Positron emission tomography and dynamic contrast-enhanced computerised tomography interobserver variability: site versus core read

### *Positron emission tomography*

Nuclear medicine specialists or dual-trained radiologists read the PET/CT examinations. For the SUV, the ROI was placed using the threshold technique, as in standard practice.

Originally, 10% ( $n = 41$  to account for potential losses due to missing data) of the total patients undergoing PET were selected for a second read by a core laboratory. Owing to high variability

between the recruiting sites, documented results and the core laboratory read results, a full core read of all the PET data sets was performed. The original site reads were performed by the reporting physician at each of the study centres. Image analysis was performed using the onsite software used in routine clinical practice. All core laboratory reads were performed by a single radiologist who was not involved in the primary reads. The core laboratory read was blinded to the original diagnostic CT, the DCE-CT and the original site's read. A ROI was defined based on the nodule size and location, with alterations made when there was a discordance in the location of the PET uptake relative to the nodule on the CT that could be attributed to motion/respiration. Quantitative assessment of the nodule uptake was performed using the maximum standardised uptake value (SUVmax) and the mean standardised uptake value (SUVmean). The nodule CT and PET characteristics were semiquantitatively graded based on visual inspection, as described in *Table 3*.

A total of 354 cases were included in the reproducibility analysis for the SUVmax, with six excluded because of lack of an identifiable nodule, multiple nodules or missing data. The SUVmax demonstrated a small, but significant, bias between the site read and core laboratory read, and high variability between the two reads with wide limits of agreement (LOAs) (mean difference 0.37, LOA -3.36 to 4.11; paired *t*-test  $p < 0.001$ ). Despite this, when considering a threshold for the SUVmax of  $> 2.5$  as being malignant, there was excellent agreement between the site read and core laboratory read [ $\kappa$  (unweighted): 0.93;  $p < 0.001$ ].

As the SUVmax LOAs were wider than expected, we further reviewed the steps in the scanning and analysis. Several sites were identified to be using PSF reconstruction algorithms, which can result in an increase in SUVmax of up to 25%. Data submitted to the core laboratory were all reconstructed using the standard algorithms, but the reporting sites would have had access to both; therefore, it is possible that the values from the PSF reconstruction would have been recorded. When the 108 cases from those sites were excluded, leaving 247 cases in the analysis, the mean difference was -0.06 (LOAs -2.29 to 2.42; paired *t*-test  $p = 0.41$ ), producing significantly improved LOAs and the loss of bias.

*Figure 6* contains the Bland-Altman plots for the site and core laboratory reads for the SUVmax, with and without PSF sites included.

A total of 339 cases were included in the reproducibility analysis for the SUVmean, with 15 cases excluded because of lack of documentation of site read of the SUVmean. There was a small, but significant, bias in the SUVmean between the site read and the core laboratory read, with wider LOAs than those for the SUVmax (SUVmean: mean difference -0.38, LOAs -4.41 to 3.67; paired *t*-test  $p < 0.001$ ).

**TABLE 3** Semiquantitative grading scheme for both the PET and CT component of the PET/CT examination combining the anatomical features from the CT and the metabolic features from the PET exam

Grade	Significance	PET/CT	Attenuation correction CT
0	No evidence of malignancy	No visible uptake	Round, well-defined lesion with laminated or popcorn calcification
1	Low probability of malignancy	Uptake less than mediastinal blood pool	Inflammatory features, for example air bronchograms, enfolded lung
2	Indeterminate	Uptake comparable to mediastinal blood pool	Smooth well-defined margins, uniform density
3	High probability of malignancy	Uptake greater than mediastinal blood pool	Lobulated, spiculated or irregular margins
4	Very high probability of malignancy	Evidence of distant metastases (i.e. M1 disease)	Evidence of distant metastases (i.e. M1 disease)

M, metastasis.

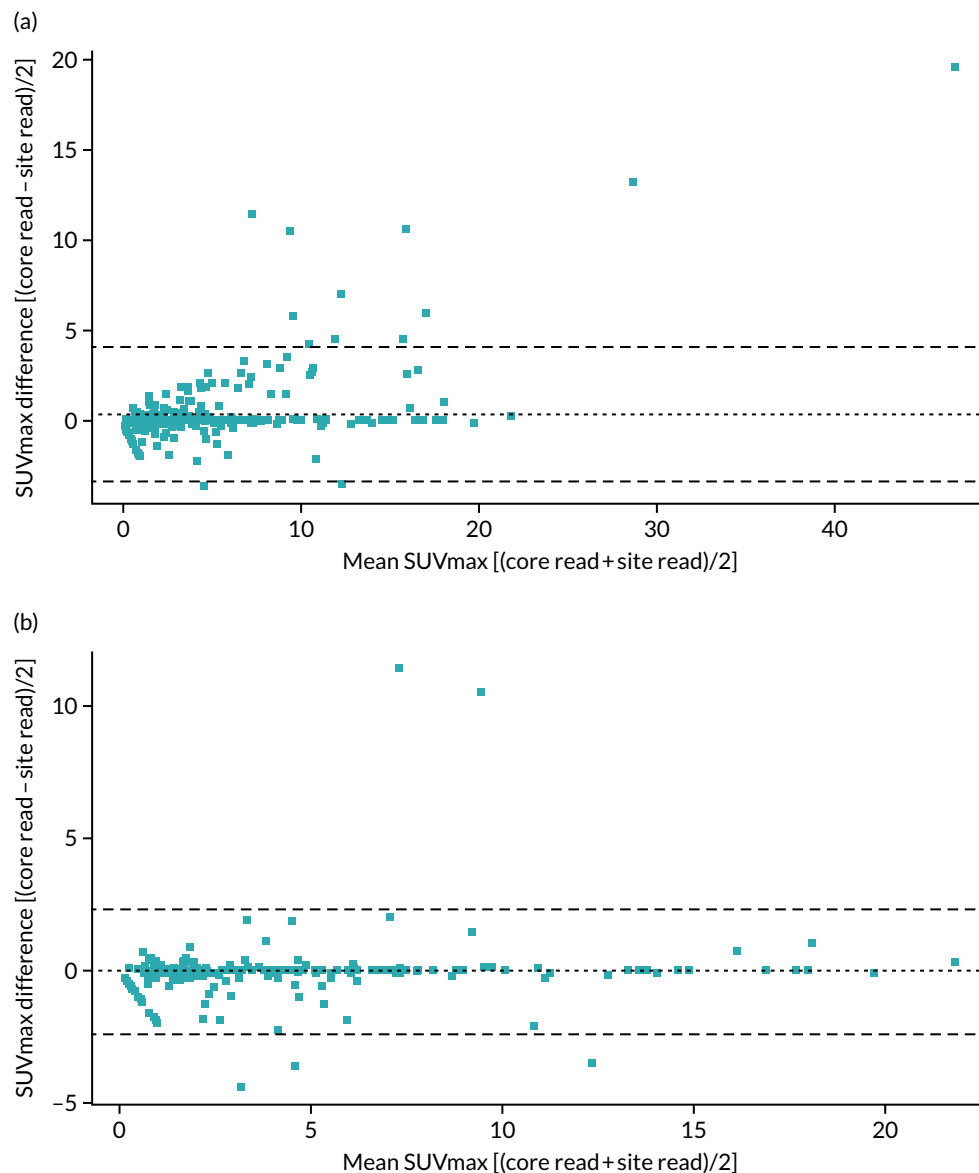


FIGURE 6 Bland–Altman plot comparing the SUVmax obtained at the recruiting site with that obtained at the core laboratory. (a) Across all sites; and (b) excluding those with PSF.

Intraobserver variability was also examined, using two different software packages [Xeleris™ (GE Healthcare, Chicago, IL, USA) and ADW 4.4 (GE Healthcare)], with a minimum of 4 months between reads to minimise the chances of retention between scans. This showed high levels of agreement for the SUVmax (mean difference  $-0.04$ , LOAs  $-0.28$  to  $-0.21$ ; paired  $t$ -test  $p = 0.07$ ), and reasonable agreement for the SUVmean, with higher variability (mean difference  $0.34$ , LOA  $-1.81$  to  $-2.49$ ; paired  $t$ -test  $p = 0.06$ ). The two outliers in both cases were nodules that lay close to the heart or to the diaphragm, making accurate measurement challenging because of the proximity of the nodule to adjacent structures, and, therefore, prone to variability.

There was excellent agreement between the sites on the visual semiquantitative grading of the PET uptake [ $\kappa$  (with squared weighting):  $0.87$ ;  $p < 0.001$ ]. The greatest variation in scoring occurred around grade 2, which is ‘uptake equivalent to the mediastinum’. Only eight (2.3%) cases were called as equivalent to the mediastinum at the sites, whereas 41 (11.6%) cases were called as equivalent to the mediastinum at core laboratory read. Given that the SUVmax agreement was tight in and around a SUVmax of 2–2.5, which is typically equivalent to the mediastinal blood pool, this suggests that it is very subjective as to when the visual uptake becomes greater, or less, than the mediastinum.

There was fair agreement in the CT visual grading of the nodules [ $\kappa$  (with squared weighting): 0.33;  $p < 0.001$ ], with disagreement between site read and core read most pronounced for grade 2 lesions. As the original CT was blinded to the core laboratory, but not to the site readers, it may be that subtle spiculation, lobulation or heterogeneity of attenuation not evident on the low-dose attenuation-corrected CT scan may, in fact, have been determined by the standard CT scan, causing this marked upgrade in scoring. Alternatively, it may be that determining whether there is true spiculation or simply background parenchymal lung disease causing contour irregularity is a poorly reproducible finding on CT.

The combined  $^{18}\text{F}$ -FDG-PET/CT assessment was classified as positive for malignancy if one of the following criteria was met:

- grade 4 on  $^{18}\text{F}$ -FDG-PET/CT
- at least grade 3 on both PET and CT appearances
- grade 2 on PET and grade 3 or 4 on CT.

Based on this grading, there was good overall agreement on the presence of malignancy [ $\kappa$  (unweighted): 0.70,  $p < 0.001$ ].

Based on the above, the SUVmax appears to be the most robust measurement across sites. It has both a lower mean difference and narrower LOAs than the SUVmean. Using a SUVmax cut-off point of 2.5 yielded significantly higher agreement between sites and the core read for malignancy than visual grading of the lesions. CT grading was particularly prone to high observer variability, which affected the diagnosis of malignancy in a large number of cases.

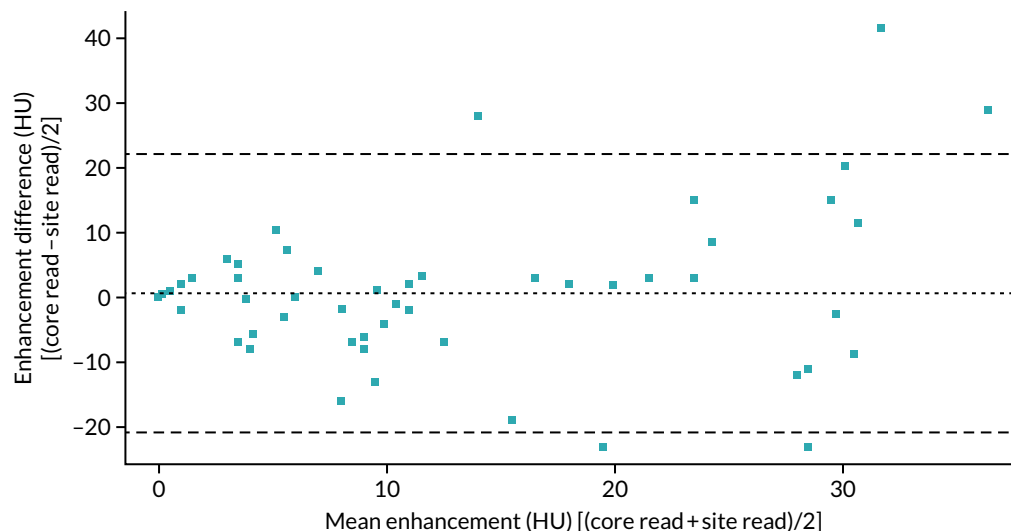
### ***Dynamic contrast-enhanced computerised tomography***

Thoracic radiologists or dual-trained radiologists read the DCE-CT. Training was given using a manual and a short video. The ROI was placed using mediastinal settings (window width: 400 HU, window level: 40 HU) in the axial plane with the largest diameter of the nodule. The two-dimensional size was measured in millimetres by taking the longest diameter of the lesion and the perpendicular diameter in the axial plane.

Twenty per cent ( $n = 66$  to account for potential losses due to missing data) of the total DCE-CT scans were selected for a second read. A larger number of DCE-CT scans were chosen to reflect the fact that DCE-CT was not generally used in routine practice, and thus was potentially more prone to variability in the analysis. A training video was prepared to ensure that site readers were analysing the DCE-CT scans using a similar technique, with the technique used in the study as previously published.<sup>45</sup> The selection criteria were weighted to ensure that all centres had at least one scan second read and that the number of scans selected from each centre was representative of the proportion of the total scans performed at that centre. The individual scans from each centre were randomly selected. The site reads were performed using the on-site software used in routine clinical practice. All central reads were performed on a single software platform (*syngo.via*, Siemens Healthineers AG, Erlangen, Germany) by a single reader. The second reader was blinded to the original diagnostic CT, the PET/CT and the original site read. Quantitative assessment of the nodule enhancement was performed using maximum enhancement.

Maximum enhancement was comparable in 63 of the 66 cases (three cases had missing data). There was no significant difference between the site read and the core laboratory read (mean difference 2.57 HU, LOAs -34.5 to 39.6; paired  $t$ -test  $p = 0.29$ ). The Bland-Altman plot for maximum enhancement is shown in *Figure 7*.

Considering a threshold of maximum enhancement of  $\geq 20$  HU as being malignant, there was good agreement between the site read and the core laboratory read [ $\kappa$  (unweighted): 0.75;  $p < 0.001$ ].



**FIGURE 7** Bland-Altman plot comparing maximum enhancement obtained at the recruiting site with that obtained at the core laboratory.

There was no systematic bias in the values obtained by the DCE-CT central read and the sites in either the maximum enhancement of the nodules or the wash-out. There were, however, wide LOAs between the two measures, suggesting substantial variability in the technique. Further work is required to determine if this is due to differences in analysis software/scanners, or due to variations in the size and precise location of each of the ROIs at each of the time points.



# Chapter 4 Diagnostic accuracy systematic review and meta-analysis

## Methods

The study was prospectively enrolled in PROSPERO (CRD42018112215). The study has been reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses of Diagnostic Test Accuracy Studies (PRISMA-DTA) statement.<sup>46</sup>

The condition to be studied was the diagnostic test accuracy of SPNs. The inclusion criterion was studies examining SPNs being worked up for malignancy; studies that included participants aged < 18 years and studies of pure ground glass nodules were excluded. The intervention of interest was DCE-CT. CT scans were included as long as there was a minimum of both a pre-contrast-enhanced and post-contrast-enhanced CT data set for the quantification of the degree of enhancement. The gold standard against which the test was examined was required to be histological diagnosis of malignancy obtained from either needle biopsy or surgical resection, with benign status confirmed either histologically, or with follow-up imaging showing no growth at 2 years or resolution. We considered both prospective and retrospective diagnostic accuracy studies that contained sufficient data to construct contingency tables in order to assess true-positive, false-positive, true-negative and false-negative results.

To identify articles of interest for review, MEDLINE and EMBASE were searched from their inception until October 2018 for published studies on the diagnostic accuracy of DCE-CT in the characterisation of pulmonary nodules. The full search strategy is documented in *Appendix 5*. Titles and abstracts of studies retrieved using the search strategy, and those from additional sources, were all independently screened by two reviewers (Jonathan R Weir-McCall and Stella Joyce) to identify studies that potentially met the inclusion criterion. The full texts of these potentially eligible studies were retrieved and independently reviewed by the two reviewers to assess eligibility. When there was a disagreement between the reviewers, a consensus was reached through discussion. The references of the retrieved full-text articles were screened for further articles of interest; if any articles were found, these were retrieved if they had not been identified by the original search strategy.

A single reviewer (Jonathan R Weir-McCall) used a standardised, pre-piloted form to extract data from the included studies for assessment of study quality and evidence synthesis. Extracted information included study population, participant demographics and baseline characteristics; details of the CT scanning hardware, scanning technique and diagnostic threshold used; study methodology; nodule size range and eventual diagnosis; diagnostic accuracy metrics; and radiation dose. Two reviewers authors (Jonathan R Weir-McCall and Stella Joyce) independently assessed the risk of bias in the included studies through the use of the second version of the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) questionnaire.<sup>47</sup> Discordance in the scoring of bias between the two reviewers was resolved by a third reviewer (Lena-Marie Dendl).

Two deviations occurred from the original pre-registered protocol. A size threshold was not prespecified in the original protocol, yet, during the literature review, it became readily apparent that the upper size limit included in studies varied markedly. Although the Fleischner<sup>9</sup> and BTS<sup>8</sup> guidelines state that the upper limit of a SPN size is 30 mm, we allowed up to 40 mm for the purpose of this analysis because of the high quality of many of the studies using this threshold, and the granularity it would provide the review. An analysis was performed to compare studies with nodules of > 30 mm with studies with nodules of  $\leq$  30 mm, as described in *Statistical analysis*, to determine the effect this might have on the results. The original protocol called for the analysis of SPNs; several studies recruited

based on the detection of a SPN, but, if an additional nodule was detected at the time of the index test, they included, analysed and followed up both lesions. Despite not being SPN studies, these were included in the analysis as they reflect routine clinical practice whereby new lesions can frequently be detected on interval studies, or picked up when CT is performed following detection of a nodule on chest radiographs.

## Statistical analysis

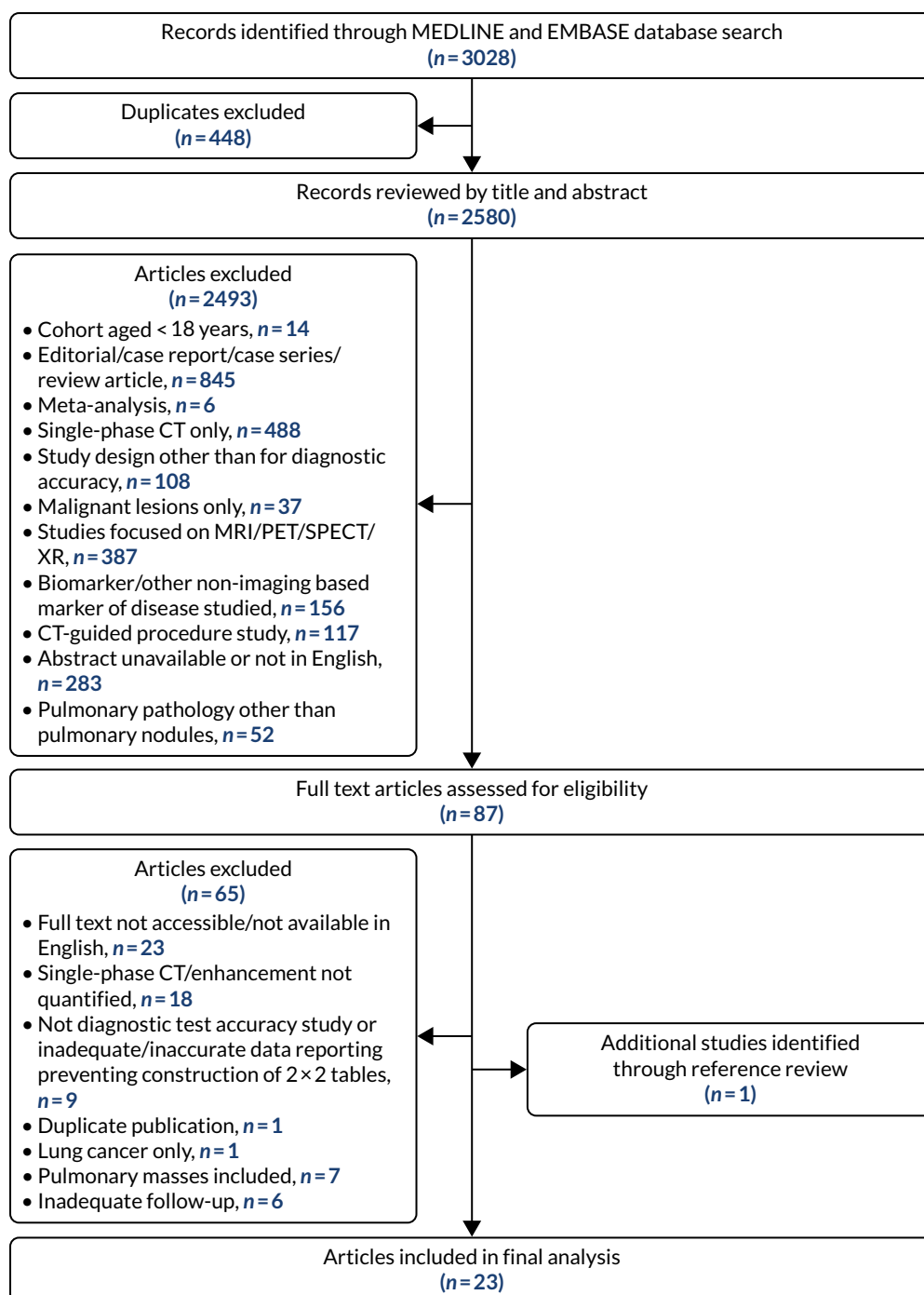
Numbers of true positives, false positives, true negatives and false negatives were extracted from the studies and used to form  $2 \times 2$  contingency tables, which were used to derive sensitivity, specificity, positive likelihood ratio (PLR), negative likelihood ratio (NLR) and diagnostic odds ratio (DOR). Results were pooled using the lme4 package in R (Rstudio, version 1.1.463, RStudio, Inc. Boston, MA, USA; R, The R Foundation for Statistical Computing, Vienna, Austria) to perform a bivariate binomial random-effects meta-analysis.<sup>48</sup> This uses a binary (logit) generalised linear mixed-model fit by maximum likelihood (using a Laplace approximation). Bivariate summary receiver operating characteristic (SROC) curves were constructed using the bivariate random-effects model outputs to populate the SROC plot in Review Manager version 5.3. (RevMan, The Cochrane Collaboration, The Nordic Cochrane Centre, Copenhagen, Denmark). To identify potential sources of heterogeneity, we stratified a secondary analysis into subgroups according to characteristics such as sample size, lesion size, risk of bias (low vs. high/indeterminate), diagnostic thresholds and whether or not the diagnostic threshold was prospectively set. These were included as covariates, in turn, in a meta-regression analysis, with analysis of statistical significance between models performed using a likelihood ratio test of nested models. For sample size, the threshold at which to split the data was arbitrarily set at 100, to represent larger samples that were less likely to be prone to bias due to outliers. For mean nodule size, the sample was split at 20 mm, to provide a reasonable split of the data. For maximum nodule size, the data were split based on whether or not the study included nodules of  $> 30$  mm, as the 30-mm diameter is considered by most guidelines as the upper threshold for a lesion to be called a nodule, after which it is considered to be a mass. The effect of publication date was examined by splitting on the median (2008), with studies published in the previous decade considered to be more representative of modern CT technology. In studies reporting the diagnostic accuracy of multiple thresholds, the optimal threshold was used in the primary analysis.

In the secondary analyses examining different thresholds, studies were included in each subgroup analysis if they had reported the threshold of interest. Thresholds with one or two studies reporting the same threshold were not considered for meta-analysis. To test for study publication bias and heterogeneity, a Galbraith plot was created to examine the interaction between the efficient score and variance, with the Harbord test used to test for funnel plot asymmetry.<sup>49</sup> All statistical analyses were performed using RStudio version 1.1.463. Forest plots and SROC curves were generated using RevMan version 5.

## Results

Of 3028 potential papers identified by the literature review, 22 met the inclusion criterion. An additional study was located from the references of the included papers, resulting in 23 studies in the final analysis. *Figure 8* details the study flow diagram of the studies identified and screened for eligibility, and the reasons for study exclusion.

The 23 included studies incorporated data from 2397 patients with 2514 nodules. Out of 2514 nodules, 1389 (55.3%) were malignant. The studies were predominantly retrospective single-centre studies. *Appendix 5, Tables 31 and 32*, details the study characteristics of each of these studies and the scanning technique, injection protocol and reconstruction algorithm used in each.



**FIGURE 8** The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram of the papers identified by the literature search, screened for eligibility and included in the final study. MRI, magnetic resonance imaging; SPECT, single-photon emission computerised tomography; XR, X-ray.

The results of the QUADAS-2 bias and applicability assessment are summarised in *Figure 9. Appendix 5, Table 33*, documents the individual bias scores for the seven domains for all included studies. Bias in patient selection was unclear in a large number of studies [14/23 (61%)] because of a lack of reporting of the sampling of patients for the diagnostic test accuracy evaluation, with many retrospective studies not clearly documenting whether or not consecutive cases were included. Risk of bias in the index test was high in a large number of studies [12/23 (52%)] because of a lack of prespecification of the intended threshold to be used, and, in several studies, multiple techniques of enhancement of quantification were used simultaneously (including, but not limited to, absolute contrast enhancement,

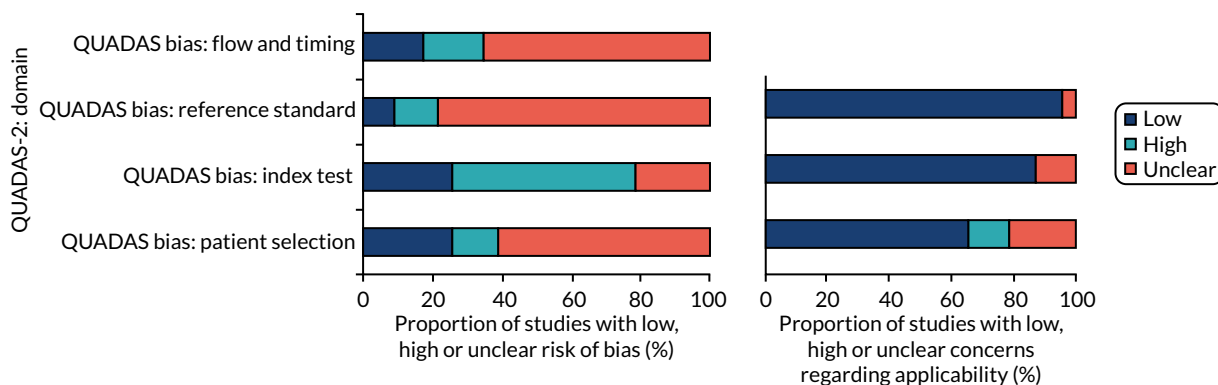


FIGURE 9 The QUADAS-2 scoring summary of the included studies.

relative contrast enhancement, wash-in, wash-out, wash-in and wash-out, and area under the enhancement curve). Bias regarding the reference standard was unclear in the majority of studies [18/23 (78%)], with the blinding of the reference standard to the index test infrequently reported. Flow and timing had a similarly high-rate frequency of uncertain bias [15/23 (65%)], with the delay between the index test and reference standard infrequently reported. Concerns regarding the applicability of the included studies to the review question were low for the majority of the studies (see Figure 9).

The results of the individual studies' sensitivities and specificities are collated in a forest plot in Appendix 5, Figure 24, with all studies reporting a per-nodule diagnostic accuracy. The pooled analysis of the 23 studies is reported in Appendix 5, Table 34. The pooled sensitivity and specificity were 94.8% (95% CI 91.5% to 96.9%) and 75.5% (95% CI 69.4% to 80.6%), respectively (the SROC plot is presented in Figure 10), with a PLR of 3.86 (95% CI 2.99 to 4.74), a NLR of 0.07 (95% CI 0.03 to 0.10) and a DOR of 56.6 (95% CI 24.2 to 88.9). Only two distinct enhancement thresholds were reported by more than two studies, with the pooled analysis for each of these reported in Appendix 5, Table 34. Of these, a threshold of < 20 HU enhancement for the differentiation of a malignant from a benign nodule had the highest DOR of 142.5 (95% CI -36.4 to 321.3), maintaining a high sensitivity of 98.3% (95% CI 95.1% to 99.4%) and moderate specificity of 71.0% (95% CI 63.1% to 77.8%). However, it should be noted

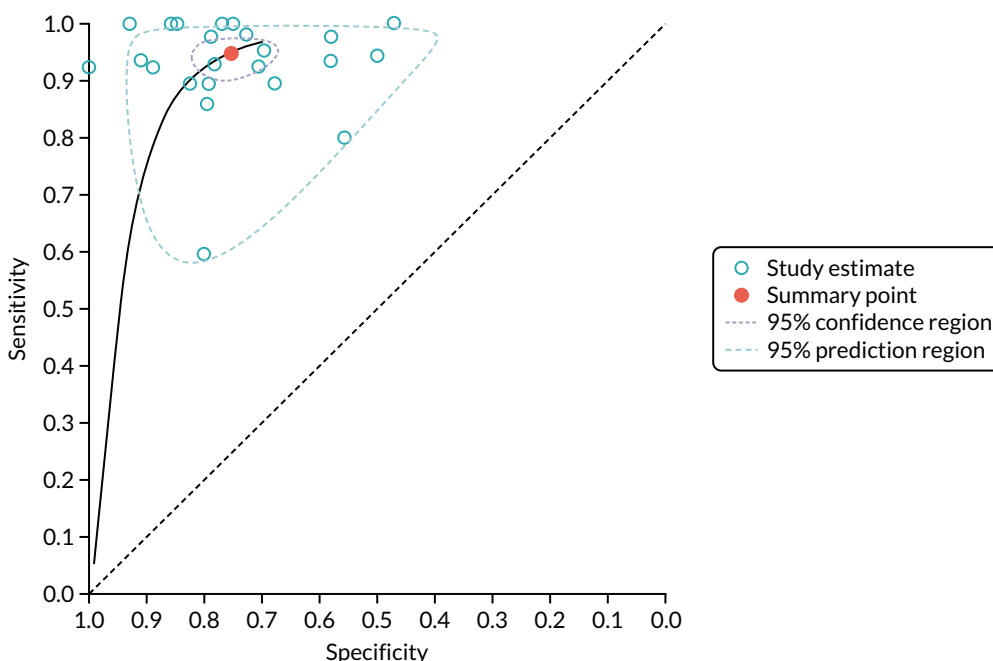


FIGURE 10 Bivariate SROC curve of the included studies.

that the CIs were wide for this threshold, as there were too few data points, with far narrower CIs for a 15-HU threshold (see *Appendix 5, Table 34*).

The aqua circles indicate individual studies and the orange circle indicates the summary point. The purple dashed line is the 95% confidence region for the summary operating point, whereas the light-blue dashed line is the 95% prediction region (which is the confidence region for a forecast of the true sensitivity and specificity in any future study).

The subgroup analyses for heterogeneity studies with a low risk of reference standard bias demonstrated a borderline higher sensitivity and lower specificity than studies with an intermediate/high risk of reference standard bias ( $p = 0.044$ ). No difference was present between subgroups when studies were split based on CT technique; sample size; mean or maximum nodule size; threshold prospectively or retrospectively set; or the presence of patient selection bias, index test bias or flow and timing bias ( $p > 0.1$  for all, full results are described in *Appendix 5, Table 35*). In particular, there were no significant differences in the pooled sensitivity or specificity between studies that only included nodules of  $\leq 30$  mm (and therefore met the current definition of SPN) and studies that included larger nodules of up to 40 mm in size ( $p = 0.07$  for between-group differences in sensitivity and specificity).

The Galbraith plot (see *Appendix 5, Figure 25*) demonstrated multiple studies falling outwith the 95% CIs, consistent with a significant interstudy heterogeneity in findings, but there was not any significant asymmetry in the plot ( $p = 0.87$ ) to suggest publication bias.

## Discussion

This meta-analysis demonstrates a high sensitivity and moderate specificity for DCE-CT for the diagnosis of SPNs, with a pooled sensitivity and specificity of 94.8% and 75.5%, respectively. However, the study quality was indeterminate in a significant number of the studies, with only one multicentre study and a large number of small studies. Although the analysis shows promising results for the technique, the low quality of the included studies must be taken into account; further carefully designed high-quality multicentre studies are required.

The current Fleischner guidelines<sup>9</sup> for further investigation and management of indeterminate SPNs call for either PET/CT or biopsy if the nodule is  $> 8$  mm in size; DCE-CT is not mentioned in the diagnostic pathway, despite inclusion of the technique in the 2005 version of the guidelines.<sup>10</sup> The BTS guidelines<sup>8</sup> state that DCE-CT should not be used when PET is available, although it is acknowledged that there is little evidence to support this beyond the historical prerogative of PET/CT. A 2018 meta-analysis<sup>12</sup> of PET/CT including 20 studies with 1557 participants reported a sensitivity and specificity of 89% and 70%, respectively, and a DOR of 22. These results are similar to the DCE-CT results obtained in this meta-analysis: 23 studies including 2397 participants demonstrated a pooled sensitivity, specificity and DOR of 95%, 76% and 57, respectively. This suggests that DCE-CT could replace PET/CT as an equivalent diagnostic technique. A formal comparison of the two techniques is recommended to confirm this, as there are a limited number of studies directly comparing DCE-CT with PET/CT. The limited number of studies currently precludes the ability to perform a meta-analytic comparison. Ohno *et al.*<sup>33</sup> compared DCE-CT with both PET/CT and dynamic contrast-enhanced magnetic resonance imaging (MRI) in a single-centre study of 198 patients, and found that DCE-CT outperformed both MRI and PET/CT in specificity and accuracy. This contradicted the results of Yi *et al.*,<sup>17</sup> who found, in a single-centre study of 119 participants, that PET/CT was more sensitive than DCE-CT, with specificity equal to that of DCE-CT. Thus, further work is required to directly compare these two modalities. Another technique that has a growing body of evidence is that of diffusion-weighted MRI. Whereas PET/CT examines metabolism and DCE-CT measures perfusion, diffusion-weighted MRI quantifies the movement of water within the lesion. A 2019 meta-analysis<sup>50</sup> of diffusion-weighted

MRI for the diagnosis of indeterminate SPNs has suggested superiority of this technique over PET/CT, with a pooled sensitivity, specificity and DOR of 83%, 91% and 50, respectively, for diffusion-weighted MRI, compared with 78%, 81% and 15, respectively, for PET/CT.<sup>50</sup> Given the differing nature of the three parameters in question, further research is needed, both to compare them and to determine whether the information from perfusion, diffusion and metabolism are complementary or duplicative in improving diagnostic accuracy.

The equivalent sensitivity, specificity and accuracy in this meta-analysis of DCE-CT, compared with previous meta-analyses of PET/CT, provide supportive evidence for consideration of the incorporation of DCE-CT into the diagnostic pathway of pulmonary nodules. CT machines are both more commonly found and more readily accessible in hospital settings than PET/CT equipment. A dynamic contrast examination is very similar to a standard contrast CT procedure, which is commonly undertaken at all hospitals and requires no additional equipment. A PET/CT examination requires the injection of a radioactive substrate, which needs to be delivered reliably to centres undertaking PET examinations. The requirement of such a supply chain can have significant impact on service flexibility and can result in imaging cancellations when there is disruption or delay in the delivery of the radioactive agent.<sup>51</sup> Future studies examining whether or not certain subgroups of pulmonary nodules (such as those of small size), or nodules found in patients with different risk profiles and likelihood of malignancy, may have more to gain from a DCE-CT examination than from a PET/CT procedure are also required. Similarly, a tiered approach using DCE-CT as the first diagnostic test and using the result as a gatekeeper, with PET/CT as the follow-on examination, may allow for a more considered, nuanced approach targeting a more appropriate workup, utilising the strengths of both techniques. Indeed, such an approach has been shown to be a cost-effective approach for the diagnosis of SPNs.<sup>26</sup> Robust direct comparative accuracy studies of DCE-CT and PET/CT in the same population, and cost-effectiveness studies are warranted to test the various diagnostic pathways.

There are several limitations with the current meta-analysis. The quality of the included studies was frequently indeterminable because of a lack of reporting of key metrics. The studies were almost exclusively single centre, and frequently retrospective, both of which are likely to amplify the apparent diagnostic accuracy of the technique. In addition, the dynamic contrast acquisition technique and the metrics for the quantification of the enhancement were heterogeneous throughout the studies. Although these factors did not appear to have an impact on the accuracy of meta-regression, a standardised acquisition and analysis technique should be agreed upon to improve reproducibility, to facilitate comparison between trials and to allow more widespread adoption.

In conclusion, we have found a high diagnostic accuracy of DCE-CT for the diagnosis of pulmonary nodules, although study quality was poor or indeterminate in a large number of cases. These findings support the current investigation into how DCE-CT may complement or augment the current diagnostic pathway of pulmonary nodules.

# Chapter 5 Systematic review of cost-effectiveness

## Methods

The methods for the systematic review were outlined in a research protocol prior to undertaking the review; the review was registered on the PROSPERO database (CRD42019124299). The systematic review aimed to identify economic evaluations of diagnostic imaging techniques for characterising SPNs. The review was not limited to studies that included PET, with or without integrated CT, and DCE-CT in their diagnostic strategies, as the review had broader aims:

- to identify and summarise evidence on the cost-effectiveness of alternative imaging modalities and combinations of imaging modalities for the characterisation of SPNs, with a particular focus on the relevance of such evidence for decision-making in the NHS
- to identify methods used in the cost-effectiveness studies (studies using model-based synthesis, data from observational studies or a combination) and the range of outcomes reported.

### Search strategy

We used a search strategy based on SPN (keyword and text), diagnostic imaging technology and cost-effectiveness terms (the full MEDLINE search strategy is listed in *Appendix 6*) to search the following electronic databases: MEDLINE (including In-Process & Other Non-Indexed Citations), EMBASE, Web of Science, Bioscience Information Service (BIOSIS), The Cochrane Library (HTA database, NHS Economic Evaluation Database), the Centre for Reviews and Dissemination (HTA, Database of Abstracts of Reviews of Effects, NHS Economic Evaluation Database) and Science Direct. Reference lists of relevant reviews identified in our searches, and key papers identified by subject matter experts in the SPutNik study group, were searched for additional references. Initial searches were conducted in July 2013 and searched the databases from inception to that date. Updated searches, using the same databases and identical search strategies, were conducted in September 2015 and November 2018.

### Study selection

*Table 4* reports the inclusion criteria for the systematic review. For the initial screen, titles and abstracts of studies identified by the search strategy were assessed for potential eligibility by two reviewers independently. In cases when neither reviewer was a health economist, inclusion decisions were checked by a health economist. Papers included at the initial screen were retrieved, with the full text screened independently by two reviewers. Differences in opinion were resolved at each stage through discussion.

TABLE 4 Inclusion criteria for systematic review of economic evaluations

Grouping	Characteristic
Population	Patients under investigation for a SPN
Intervention	Strategies involving CT or PET
Comparator	Resection, biopsy and/or clinical follow-up
Setting	Secondary or tertiary care
Outcomes	Costs, cost per case detected and incremental cost per life-year/QALY gained
Design	Cost, cost-effectiveness and cost-utility studies

**Data extraction, critical appraisal and synthesis**

Data extraction and quality assessment were undertaken by one reviewer, and were checked by a second reviewer. Differences in opinion were resolved at each stage through discussion.

A standard template was developed for data extraction on study characteristics, data inputs used in the economic analysis [specifically, all diagnostic accuracy estimates (including, when relevant, methods used for searching for and pooling diagnostic accuracy data from other studies), costs and health-state utility], approaches to/results from the sensitivity analysis, and cost-effectiveness results. Data were synthesised through a narrative review, with outcomes presented as incremental cost-accuracy ratios or ICERs.

Studies were critically appraised using an adapted checklist (see *Appendix 7*) based on Drummond *et al.*<sup>52</sup> and Philips *et al.*<sup>53</sup> The checklist was used to assess the methodological quality of studies in terms of data inputs, assumptions, model structure and presentation of results. Responses to items in the critical appraisal checklist were tabulated and discussed in narrative form, with no summary score or overall statement (e.g. high or low quality).<sup>54</sup> Studies were also assessed for compliance with current methodological guidance applying to the appraisal of diagnostic assessments in NHS decision-making.<sup>11</sup>

**Results of systematic review of economic evaluations**

The original search identified 382 candidate publications, of which 362 were excluded by screening titles and abstracts. Of the 20 papers retrieved for full screening, 12 did not meet the inclusion criteria (*Figure 11*) (a list of relevant excluded studies can be found in *Appendix 8*). Eight studies met all inclusion criteria: Gambhir *et al.*,<sup>22</sup> Dietlein *et al.*,<sup>24</sup> Keith *et al.*,<sup>21</sup> Comber *et al.*,<sup>26</sup> Gould *et al.*,<sup>20</sup> Gugiatti *et al.*,<sup>25</sup> Tsushima and Endo<sup>27</sup> and Lejeune *et al.*<sup>28</sup> The first updated search identified 98 candidate publications, of which 90 were excluded by screening the titles and abstracts. Of the eight papers retrieved for full screening, seven did not meet the inclusion criteria (see *Figure 11*, and see *Appendix 8* for a list of relevant excluded studies). One study met all inclusion criteria: Deppen *et al.*<sup>55</sup> The second updated search identified 184 candidate publications, with all 184 excluded after screening titles and abstracts. As a result, nine papers were included in the review.

**Study characteristics**

The key characteristics of the included studies are presented in *Table 5*. The included studies were conducted in several countries [USA ( $n = 3$ ),<sup>20,22,55</sup> Australia ( $n = 2$ ),<sup>21,26</sup> Germany ( $n = 1$ ),<sup>24</sup> Italy ( $n = 1$ ),<sup>25</sup> Japan ( $n = 1$ )<sup>27</sup> and France ( $n = 1$ )<sup>28</sup>], with costs in the studies using a range of currencies [US dollars ( $n = 3$ ),<sup>20,22,55</sup> Australian dollars ( $n = 2$ ),<sup>21,26</sup> euros ( $n = 4$ )<sup>24,25,28</sup> and yen ( $n = 1$ )<sup>27</sup>] and reporting different base years (ranging from 1995 to 2011, when stated). The perspectives adopted by the studies ranged from a societal<sup>20</sup> to a health service perspective.<sup>24,25,27,28,55</sup> Four studies did not state the perspective taken.<sup>21,26,27,55</sup>

The type of economic evaluation varied, with one cost-minimisation analysis,<sup>25</sup> six cost-effectiveness analyses<sup>22,24,26-28</sup> and two cost-utility analyses<sup>20,55</sup> (see *Table 5*). All the analyses involved some form of modelling to estimate the costs of diagnostic workup/management, as well as diagnostic yield/outcome, although studies differed significantly in the extent to which they attempted to extrapolate beyond diagnostic accuracy to final patient outcomes. As a result, the studies present a diverse range of cost-effectiveness results, including cost per case, incremental cost per life-year gained and incremental cost per QALY gained.

Although all studies included people with single pulmonary nodules, the size of the nodules and the approaches used to identify them and determine their sizes were not always stated. Six studies identified that the SPN was  $\leq 4$  cm,<sup>20,22,24,27,28,55</sup> three studies did not state the size of the SPN,<sup>21,25,26</sup> and two reported that the SPN was indeterminate in size.<sup>21,26</sup> CT scans and/or X-rays were used to identify and determine the nodule size in six studies;<sup>20,22,24,27,28,55</sup> three studies failed to state the approach taken.<sup>21,25,26</sup>

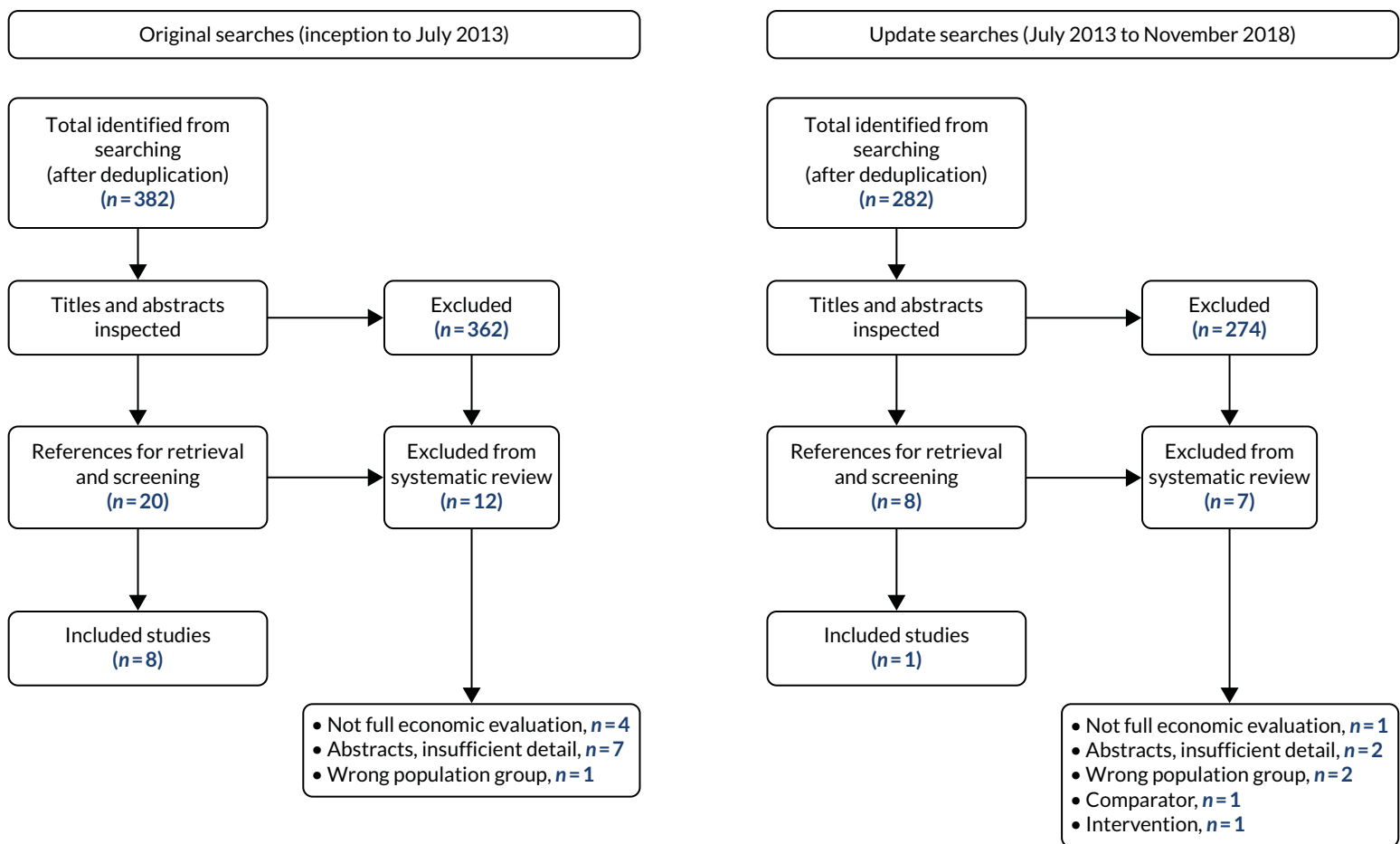


FIGURE 11 Study selection: original and updated searches.

TABLE 5 Characteristics of included studies

Study	Stated study objective	Study type	Perspective	Country	Currency (base year)	Population	Comparisons (see Table 6 for full details)	Outcomes
Gambhir <i>et al.</i> <sup>22</sup> 1998	Assess cost-effectiveness of strategies for diagnosis and management of SPN	CEA	Not stated	USA	US\$ (1995)	SPN of < 3 cm on X-ray	Four (WW, surgery, CT, CT then PET)	Cost, life-years, incremental cost per life-year gained
Dietlein <i>et al.</i> <sup>24</sup> 2000	Assess cost-effectiveness of adding PET/CT to diagnosis/characterisation of SPNs	CEA	Public health provider	Germany	Euro (1999)	SPN of < 3 cm on X-ray and CT scan	Four (WW, biopsy, surgery, PET/CT)	Cost, life-years, incremental cost per life-year gained
Keith <i>et al.</i> <sup>21</sup> 2002	Assess cost-effectiveness of adding PET/CT to diagnosis of SPNs in Australia	CEA	Not stated	Australia	AUS\$ and euro (not stated; 1999/2000 <sup>a</sup> )	Adults, indeterminate SPN (size not stated)	Two (CT, PET) using ICP model, two (CT, CT then PET) using Gambhir <i>et al.</i> <sup>22</sup> model	Accuracy, cost per case, ICAR
Comber <i>et al.</i> <sup>26</sup> 2003	Assess cost-effectiveness of adding QECT to the diagnosis of a SPN	CEA	Not stated	Australia	AUS\$ (not stated; 1999/2000 <sup>a</sup> )	Adults, indeterminate SPN (size not stated)	Four (CT, CT then PET, CT then QECT, CT then QECT then PET)	Accuracy, cost per case, ICAR
Gould <i>et al.</i> <sup>20</sup> 2003	To assess cost-effectiveness of strategies for diagnosis and management of SPNs, particularly adding PET/CT	CUA	Societal <sup>b</sup>	USA	US\$ (2001)	Adults, new non-calcified SPN on chest X-ray (2-cm diameter)	40 combinations of five diagnostic interventions: CT, PET, TNB, surgery, WW	Cost, QALYs, incremental cost per QALY gained
Gugiatti <i>et al.</i> <sup>25</sup> 2004	Assess cost-effectiveness of adding PET/CT to diagnosis of SPN in Italy	CMA	Health service	Italy	Euro (not stated)	SPN (size not stated)	Two (CT, CT then PET)	Cost per case
Tsushima and Endo <sup>27</sup> 2004	Assess cost-effectiveness of CT-guided needle biopsy and PET/CT in diagnosis of SPN in Japan	CEA	Health service	Japan	Yen (2002)	SPN of between 1 and 4 cm in diameter, identified by X-ray	Four (CT, CT then PET, CT then PET then CT-guided needle biopsy, CT then CT-guided needle biopsy)	Accuracy, cost per case, ICAR

Study	Stated study objective	Study type	Perspective	Country	Currency (base year)	Population	Comparisons (see Table 6 for full details)	Outcomes
Lejeune <i>et al.</i> <sup>28</sup> 2005	Assess cost-effectiveness of adding PET for management of SPNs in France	CEA	Health service	France	Euro (2003) <sup>c</sup>	SPN of < 3 cm on abdominal–pelvic–thoracic CT scan	Three (WW, PET/CT, CT and PET)	Cost, life-years, incremental cost per life-year gained
Deppen <i>et al.</i> <sup>55</sup> 2014	Assess cost-effectiveness of strategies for initial diagnosis and management of SPNs	CUA	Health service	USA	US\$ (2011)	SPN of 1.5 to 2 cm, detected by CT	Four (surgery, PET/CT, CT-guided fine-needle aspiration, navigation bronchoscopy)	Cost, QALYs, incremental cost per QALY gained

CEA, cost-effectiveness analysis; CMA, cost-minimisation analysis; CUA, cost-utility analysis; ICAR, incremental cost-accuracy ratio; ICP, Institute of Clinical PET; QECT, quantitative contrast-enhanced computerised tomography; TNB, transthoracic needle biopsy; WW, watchful waiting.

a Base year not specifically stated; 1999/2000 assumed because of data sources used for costs and benefits.

b States perspective as societal in abstract, but it is unclear what this means.

c Costs for lobectomy, video-assisted thoracoscopic surgery and wedge resection derived from 2003 French National Hospital Database on Diagnosis-Related Groups ('Groupe Homogene de Malades') valued using the French national public cost scale ('Echelle Nationale des Coûts') – year for costs not given. Neither source was fully referenced. Costs for diagnostic procedures (radiography, CT and TNB) were taken from a standard schedule ('Nomenclature Generale des Actes Professionals', no year or full reference given). The cost of a short inpatient stay (12–24 hours), taken from a cost schedule for the French health insurance fund ('Tarifs de Prestations Journalieres'), was added to the schedule cost for TNB. Costs for PET were based on two published studies, conducted in France (publication years 1998<sup>22</sup> and 2000<sup>24</sup>). The paper does not report whether or not costs were converted to a common base year.

The diagnostic pathways that were evaluated differed between the included studies (see *Table 5* and *Appendix 9, Table 36*). The strategies that were compared encompassed diagnostic tests, either singly or in combination, that included CT, DCE-CT, PET/CT, quantitative contrast-enhanced computerised tomography (QECT), watchful waiting, navigation bronchoscopy, biopsy and surgery. Although there was variation in terms of the specific diagnostic strategies being evaluated, studies were generally concerned with determining the improvement in diagnostic accuracy achieved by adding a more specific test, following initial CT (i.e. offering the new test for all SPNs not characterised as benign). In most cases, the additional test was PET/CT, although Comber *et al.*<sup>26</sup> also modelled the diagnostic accuracy of strategies including QECT (see *Appendix 9, Table 36*), and Gould *et al.*<sup>20</sup> included DCE-CT in a sensitivity analysis.

Although the majority of studies compared either two,<sup>21,25</sup> three<sup>28</sup> or four strategies,<sup>22,24,26,27,55</sup> Gould *et al.*<sup>20</sup> included 40 strategies (including <sup>18</sup>F-FDG-PET/CT before CT). These strategies involved combinations of five diagnostics interventions, specifically CT, PET/CT, transthoracic needle biopsy (TNB), surgery and watchful waiting. It is unclear from the paper<sup>20</sup> whether or not all of these would be deemed clinically plausible. The only restrictions applied to the definition of allowable strategies was that CT and/or <sup>18</sup>F-FDG-PET/CT would never be performed after needle biopsy or surgery, and that needle biopsy and/or observation would never follow surgery. In contrast to some studies, Gould *et al.*<sup>20</sup> treated surgery (based on outcomes of imaging tests alone) and biopsy prior to surgery as separate strategies. Other studies distributed patients between surgery (80–85%) and biopsy (15–20%) using proportions that were unrelated to the prior probability of malignancy. In *Appendix 9, Table 36*, the details of only a selection of the strategies for which Gould *et al.*<sup>20</sup> reported cost-effectiveness results are listed.

### Quality assessment

Assessment of the methodological quality identified areas where studies appeared to follow good practice in health economic modelling and identified studies for which the approach appeared less transparent and/or rigorous (see *Appendix 9, Table 37*). All studies provided a clear statement of their decision problem; used an appropriate study type and modelling methodology; described their model structure; listed and justified the assumptions about the model structure; and described, justified and valued the resource inputs appropriately. A description and justification of data inputs to the model was also provided by all studies, except that by Gugiatti *et al.*<sup>25</sup> Although eight studies produced an incremental analysis of both the costs and consequences of different strategies,<sup>20–22,24,26–28,55</sup> seven of these studies did not conduct a fully incremental analysis, focusing on a comparison with a common baseline, rather than with the next best option.<sup>21,22,24,26–28,55</sup> In addition, Gugiatti *et al.*<sup>25</sup> failed to conduct an incremental analysis.

Several components of the economic evaluations were either not reported or not undertaken, raising concerns about the quality of the studies. Seven studies did not report conducting a systematic review to establish the diagnostic accuracy of the strategies compared.<sup>21,22,24–28</sup> This included studies that did not report undertaking systematic searches, applying prespecified study selection criteria or conducting appropriate methods of data synthesis. It appeared that there was limited or no formal quality assessment of diagnostic studies against recognised guidance.<sup>56</sup> In instances when studies used a single diagnostic study, or a published meta-analysis, as a source of diagnostic accuracy estimates, no consideration was given to the appropriateness or quality of the studies as a source for model inputs.

The comparators adopted by the different studies varied, not always representing diagnostic strategies that were used in the health service. Although four studies assessed strategies that were current practice,<sup>20,25,28,55</sup> five studies evaluated comparators that were not routinely used.<sup>21,22,24,26,27</sup> Inevitably, this limits the evidence available to support decision-making in the health sector, whether by policy-makers, health professionals or patients. The perspective of the analysis within the studies, which underpins the type of costs and health benefits included in the economic evaluation, was not always clearly stated. In five studies, either the perspective was not specified or details were limited,<sup>21,22,26,27,55</sup> with only four studies providing a clear statement of their approach.<sup>20,24,25,28</sup> In terms of health benefits, only two studies used QALYs to measure benefits,<sup>20,55</sup> with one of these studies using a standardised

and validated generic instrument.<sup>20</sup> Seven studies did not measure health benefits in terms of QALYs.<sup>21,22,24–28</sup> This raises concerns around the valuing of health across the different studies and the possibility of a lack of a standard comparable approach. Six studies failed to discount both costs and outcomes,<sup>21,22,25–27,55</sup> which can cause results to be misleading, particularly when the costs are incurred early on and the benefits to health occur at a later point. Uncertainty in the economic evaluation was adequately assessed in three studies,<sup>20,22,24</sup> with the remaining six studies lacking any or sufficient (e.g. multiple deterministic sensitivity analyses) sensitivity analyses to assess the influence of the different variables or assumptions used.<sup>21,25–28,55</sup> Such analyses are an important part of any economic evaluation, allowing a judgement about the robustness of the outcomes for informing decisions.

## Methodological characteristics

### *Type of economic evaluation*

Most included studies assessed the cost-effectiveness of the different strategies, reporting incremental cost per case, cost per unit of improved accuracy or cost per life-year gained. These analyses were focused on identifying the cost-effectiveness of achieving a given goal,<sup>21,22,24,26–28</sup> specifically, accurate characterisation or appropriate management of SPNs. Two studies were cost-utility analyses,<sup>20,55</sup> and reported outcomes in terms of QALYs. Only one study restricted its analysis to a cost-minimisation study, stating outcomes as cost per case.<sup>25</sup>

### *Perspective and time horizon*

The perspective adopted by most of the studies was that of the health service or third-party payer for costs, limiting outcomes to either diagnostic accuracy measures or direct health effects for patients.<sup>24,25,27,28,55</sup> This perspective is appropriate for supporting decisions that aim to maximise health gain from available health-care resources. However, as already noted, the limited outcomes considered in the studies means that these studies would largely be restricted to aiding decision-making with respect to the characterisation or management of SPNs only. One study stated that it had adopted a societal perspective, which, as the broadest perspective, should encompass a wide range of social opportunity costs related to the interventions.<sup>20</sup> Unfortunately, few details are provided regarding exactly what additional costs or outcomes are included in the analysis, providing limited support for this statement. The remaining three studies do not explicitly state the perspective of their economic evaluations.<sup>21,22,26</sup>

The time horizons over which the costs and health outcomes were assessed differed between the included studies. Although two studies adopted a lifetime horizon,<sup>20,55</sup> which tends to be recommended,<sup>40</sup> two studies used short durations of either 1 year<sup>28</sup> or 2 years;<sup>21</sup> and five studies did not report the time horizon.<sup>22,24–27</sup>

### *Decision model and natural history*

All studies used tree structures to evaluate their primary decision between diagnostic strategies, as well as to model ODA of each strategy (*Table 6*). In the case of four studies, this represented the extent of their analysis, with estimates limited to cost per case detected or cost per correct diagnosis (either true positive or true negative).<sup>21,25–27</sup> Of the remaining five studies, four used life expectancy estimates determined outside the model,<sup>22,24,28,55</sup> and one incorporated a Markov approach to model the impact of the diagnosis and management of a SPN on survival.<sup>20</sup>

In addition to the structural differences, the models differed in the manner and the extent to which they included aspects of natural history. Studies including watch and wait required assumptions regarding growth of both malignant and benign nodules.<sup>20,22,24,28</sup> Three studies assumed constant doubling of rates: Gambhir *et al.*<sup>22</sup> and Dietlein *et al.*<sup>24</sup> used a value of 90 days for the nodules to double, based on a previously published decision analysis,<sup>57</sup> whereas Gould *et al.*<sup>20</sup> used a value of 5.24 months, which was based on data from a published study.<sup>58</sup> Lejeune *et al.*<sup>28</sup> used time-related probabilities (with all malignant tumours doubling in size by 1 year), although the source for these data is unclear.

TABLE 6 Summary of model inputs in the included studies

Study	Model type	Probability of SPN being malignant	Diagnostic accuracy	Health outcome	Outcome/HRQoL/utility	Cost reported (2016 US\$)
Gambhir <i>et al.</i> <sup>22</sup> 1998	Decision tree	0.83	CT: <ul style="list-style-type: none"> <li>• Sensitivity, 0.999</li> <li>• Specificity, 0.610</li> </ul> PET: <ul style="list-style-type: none"> <li>• Sensitivity, 0.925</li> <li>• Specificity, 0.830</li> </ul> Biopsy: <ul style="list-style-type: none"> <li>• Sensitivity, 0.895</li> <li>• Specificity, 0.959</li> </ul> VATS: <ul style="list-style-type: none"> <li>• Sensitivity, 1.000</li> <li>• Specificity, 1.000</li> </ul> CXR follow-up: <ul style="list-style-type: none"> <li>• Sensitivity, NR</li> <li>• Specificity, NR</li> </ul>	Biopsy: <ul style="list-style-type: none"> <li>• Morbidity, 0.0008</li> <li>• Mortality, 0.2</li> </ul> Surgery: <sup>a</sup> <ul style="list-style-type: none"> <li>• Morbidity, 0.08</li> <li>• Mortality, 4.0 (0.5)</li> </ul> VATS: <ul style="list-style-type: none"> <li>• Morbidity, 0.0126</li> <li>• Mortality, 2.5</li> </ul>	Life expectancy (years): <ul style="list-style-type: none"> <li>• Benign, 14.8</li> <li>• Unresectable, 1.24</li> <li>• Surgical cure, 6.62</li> </ul>	<ul style="list-style-type: none"> <li>• CT: 378 (597)</li> <li>• <sup>18</sup>F-FDG-PET/CT: 1000 (1580)</li> <li>• Biopsy: 692 (1093)</li> <li>• Surgery: 14,121 (22,311)</li> <li>• VATS: 5132 (8109)</li> </ul>
Dietlein <i>et al.</i> <sup>24</sup> 2000	Decision tree	0.65	PET: <ul style="list-style-type: none"> <li>• Sensitivity, 0.95</li> <li>• Specificity, 0.80</li> </ul> CT follow-up: <ul style="list-style-type: none"> <li>• Sensitivity, NR</li> <li>• Specificity, NR</li> </ul>	Biopsy: <ul style="list-style-type: none"> <li>• Morbidity, 0.02</li> <li>• Mortality, 0.2</li> </ul> Surgery: <sup>b</sup> <ul style="list-style-type: none"> <li>• Morbidity, 0.2 (0.1)</li> <li>• Mortality, 2.9 (0.5)</li> </ul>	Life expectancy (years): <ul style="list-style-type: none"> <li>• Benign, 16.0</li> <li>• Unresectable, 1.5</li> <li>• Surgical cure, 7.0</li> </ul>	<ul style="list-style-type: none"> <li>• CT: 117.80 (174)</li> <li>• PET: 1227.10 (1813)</li> <li>• Biopsy: 1035.74 (1530)</li> <li>• Surgery: 5232.27 (7728)</li> </ul>

Study	Model type	Probability of SPN being malignant	Diagnostic accuracy	Health outcome	Outcome/HRQoL/utility	Cost reported (2016 US\$)
Keith <i>et al.</i> <sup>21</sup> 2002	Decision tree	0.54	CT: <ul style="list-style-type: none"> <li>• Sensitivity, 0.97</li> <li>• Specificity, 0.53</li> </ul> PET: <ul style="list-style-type: none"> <li>• Sensitivity, 0.93</li> <li>• Specificity, 0.96</li> </ul> CXR follow-up: <ul style="list-style-type: none"> <li>• Sensitivity, 1.00</li> <li>• Specificity, 0.90</li> </ul>	Not applicable	<ul style="list-style-type: none"> <li>• Malignant resected/benign treated conservatively = 1</li> <li>• Benign resected/malignant treated conservatively = 0</li> </ul>	<ul style="list-style-type: none"> <li>• CT: 400 (428)</li> <li>• <sup>18</sup>F-FDG-PET/CT: 1200 (1284)</li> <li>• Biopsy: 1204 (1289)</li> <li>• Surgery: 7585 (8118)</li> </ul>
Comber <i>et al.</i> <sup>26</sup> 2003	Decision tree	0.54	CT: <ul style="list-style-type: none"> <li>• Sensitivity, 0.99</li> <li>• Specificity, 0.61</li> </ul> PET: <ul style="list-style-type: none"> <li>• Sensitivity, 0.92</li> <li>• Specificity, 0.95</li> </ul> QECT: <ul style="list-style-type: none"> <li>• Sensitivity, 0.98</li> <li>• Specificity, 0.58</li> </ul> CXR follow-up: <ul style="list-style-type: none"> <li>• Sensitivity, 1.00</li> <li>• Specificity, 0.90</li> </ul>	Not applicable	See Keith <i>et al.</i> <sup>21</sup>	<ul style="list-style-type: none"> <li>• CT: 400 (428)</li> <li>• <sup>18</sup>F-FDG-PET/CT: 1200 (1284)</li> <li>• QECT: 110 (118)</li> <li>• Biopsy: 1204 (1289)</li> <li>• Surgery: 7585 (8118)</li> </ul>

continued

TABLE 6 Summary of model inputs in the included studies (continued)

Study	Model type	Probability of SPN being malignant	Diagnostic accuracy	Health outcome	Outcome/HRQoL/utility	Cost reported (2016 US\$)
Gould <i>et al.</i> <sup>20</sup> 2003	Decision tree and Markov model	0.55	CT: <ul style="list-style-type: none"> <li>• Sensitivity, 0.965</li> <li>• Specificity, 0.558</li> </ul> PET: <ul style="list-style-type: none"> <li>• Sensitivity, 0.942</li> <li>• Specificity, 0.833</li> </ul> Biopsy: <ul style="list-style-type: none"> <li>• Sensitivity, 0.963</li> <li>• Specificity, 0.980</li> </ul> DCE-CT: <sup>c</sup> <ul style="list-style-type: none"> <li>• Sensitivity, 0.98</li> <li>• Specificity, 0.58</li> </ul>	Biopsy: <ul style="list-style-type: none"> <li>• Morbidity, 0.24</li> <li>• Mortality, 0.005</li> </ul> Surgery: <ul style="list-style-type: none"> <li>• Morbidity, 0.084</li> <li>• Mortality, 4.2</li> </ul> VATS: <ul style="list-style-type: none"> <li>• Morbidity, 0.065</li> <li>• Mortality, 0.5</li> </ul>	QALYs: Malignant nodule managed by – <ul style="list-style-type: none"> <li>• Surgery, 6.57</li> <li>• Watch and wait, 5.09</li> </ul> Benign nodule managed by – <ul style="list-style-type: none"> <li>• Surgery, 11.80</li> <li>• Watch and wait, 11.85</li> </ul>	<ul style="list-style-type: none"> <li>• CT: 285 (388)</li> <li>• <sup>18</sup>F-FDG-PET/CT: 1980 (2693)</li> <li>• CT-guided needle biopsy: 583 (793)</li> <li>• Fluoroscopy-guided needle biopsy: 283 (385)</li> <li>• Surgery – lobectomy: 14,875 (20,230)</li> <li>• VATS: 11,625 (15,810)</li> </ul>
Gugiatti <i>et al.</i> <sup>25</sup> 2004	Decision tree	0.25	CT: <ul style="list-style-type: none"> <li>• Sensitivity, 0.53</li> <li>• Specificity, 0.75</li> </ul> PET: <ul style="list-style-type: none"> <li>• Sensitivity, 0.92</li> <li>• Specificity, 0.95</li> </ul>	Biopsy: <sup>d</sup> <ul style="list-style-type: none"> <li>• Morbidity, 28.0 (2)</li> <li>• Mortality, NA</li> </ul>	Not applicable	<ul style="list-style-type: none"> <li>• CT: 275.80 (438.91)</li> <li>• <sup>18</sup>F-FDG-PET/CT: 1071.65 (1705.44)</li> <li>• Biopsy: 165.25 (262.98)</li> <li>• Surgery: 2680.93 (4266.47)</li> </ul>

Study	Model type	Probability of SPN being malignant	Diagnostic accuracy	Health outcome	Outcome/HRQoL/utility	Cost reported (2016 US\$)
Tsushima and Endo <sup>27</sup> 2004	Decision tree	0.10	CT: <ul style="list-style-type: none"> <li>Sensitivity, 0.99</li> <li>Specificity, 0.63</li> </ul> PET: <ul style="list-style-type: none"> <li>Sensitivity, 0.968</li> <li>Specificity, 0.778</li> </ul> Biopsy: <ul style="list-style-type: none"> <li>Sensitivity, 0.769</li> <li>Specificity, 0.933</li> </ul> CT follow-up: <ul style="list-style-type: none"> <li>Sensitivity, 0.56</li> <li>Specificity, 0.95</li> </ul>	Not applicable	Patients correctly characterised prior to follow-up CT = 1; otherwise = 0 <sup>e</sup>	<ul style="list-style-type: none"> <li>CT: 15,420 (144)</li> <li><sup>18</sup>F-FDG-PET/CT: 80,300 (748)</li> <li>Biopsy: 70,830 (660)</li> <li>Surgery: 1,607,742 (14,984)</li> </ul>
Lejeune <i>et al.</i> <sup>28</sup> 2005	Decision tree and Markov model	0.43	CT: <ul style="list-style-type: none"> <li>Sensitivity, 0.98</li> <li>Specificity, 0.65</li> </ul> PET: <ul style="list-style-type: none"> <li>Sensitivity, 0.95</li> <li>Specificity, 0.81</li> </ul> Biopsy: <ul style="list-style-type: none"> <li>Sensitivity, 0.85</li> <li>Specificity, 0.95</li> </ul>	Biopsy: <ul style="list-style-type: none"> <li>Morbidity, 20.0</li> <li>Mortality, 0.0</li> </ul> Surgery – lobectomy: <ul style="list-style-type: none"> <li>Morbidity, 20.0</li> <li>Mortality, 3.0</li> </ul> Surgery – wedge resection: <sup>f</sup> <ul style="list-style-type: none"> <li>Morbidity, 20.0</li> <li>Mortality, 1.0 (0.5)</li> </ul>	Life expectancy at 65 years (years): <ul style="list-style-type: none"> <li>Benign, 19.2</li> <li>Surgical cure, 6.94 (4.53)</li> </ul>	<ul style="list-style-type: none"> <li>CT: 137 (180)</li> <li><sup>18</sup>F-FDG-PET/CT: 1021 (1345)</li> <li>Biopsy: 560 (736)</li> <li>Surgery: 13,947 (18,320)</li> <li>VATS and wedge resection: 8109 (10,652)</li> </ul>

continued

TABLE 6 Summary of model inputs in the included studies (continued)

Study	Model type	Probability of SPN being malignant	Diagnostic accuracy	Health outcome	Outcome/HRQoL/utility	Cost reported (2016 US\$)
Deppen <i>et al.</i> <sup>55</sup> 2014	Decision tree	0.65	PET: <ul style="list-style-type: none"> <li>• Sensitivity, 0.87</li> <li>• Specificity, 0.76</li> </ul> Biopsy: <ul style="list-style-type: none"> <li>• Sensitivity, 0.92</li> <li>• Specificity, 1.00</li> </ul> Navigation bronchoscopy: <ul style="list-style-type: none"> <li>• Sensitivity, 0.98</li> <li>• Specificity, 1.00</li> </ul> VATS: <ul style="list-style-type: none"> <li>• Sensitivity, 1.00</li> <li>• Specificity, 1.00</li> </ul>	Pneumothorax: <ul style="list-style-type: none"> <li>• Observation, 8.4</li> <li>• Tube, 6.6</li> <li>• Haemorrhage, 1.0</li> </ul> Navigation bronchoscopy - <ul style="list-style-type: none"> <li>• Observation, 1.5</li> <li>• Tube, 0.6</li> <li>• Haemorrhage, 0.0</li> </ul> VATS: <ul style="list-style-type: none"> <li>• Lobectomy -</li> <li>• Morbidity, 6.4</li> <li>• Mortality, 2.2</li> </ul> Wedge resection - <ul style="list-style-type: none"> <li>• Morbidity, 4.6</li> <li>• Mortality, 1.2</li> </ul>	QALYs: <ul style="list-style-type: none"> <li>• Benign, 20.8</li> <li>• Stage 1, 13.14</li> <li>• Stage 2, 6.06</li> <li>• Stage 3/4, 3.2</li> </ul>	<ul style="list-style-type: none"> <li>• <sup>18</sup>F-FDG-PET/CT: 1170 (1251)</li> <li>• Biopsy: 571 (611)</li> <li>• Navigation bronchoscopy: 1228 (1313)</li> </ul> VATS: <ul style="list-style-type: none"> <li>• Lobectomy - 11,493 (12,290)</li> <li>• Wedge resection - 10,944 (11,703)</li> <li>• Complication - 18,614 (19,906)</li> </ul>

CXR, chest X-ray; HRQoL, health-related quality of life; NA, not available; NR, not reported; VATS, video-assisted thoracoscopic surgery.

a Surgery mortality is 4.0 for curative and 0.5 for exploratory surgery.

b Surgery morbidity is 0.1 years for exploratory and 0.2 years for resection; surgery mortality is 0.5 exploratory and 2.9 for resection.

c Included in sensitivity analysis reported in an appendix to the paper,<sup>20</sup> using sensitivity and specificity from Swensen *et al.*<sup>15</sup>

d Biopsy morbidity is 28.0 for pneumothorax and 2 for pneumothorax with complications during X-ray CT-guided fine-needle biopsy.

e This definition does not characterise accuracy of the full diagnostic pathway [which includes follow-up CT and surgery for those SPNs that have grown (rate of growth not specified)]. Recalculating diagnostic accuracy taking follow-up CT (using sensitivity and specificity shown in the table) does not yield significantly different results for the low prevalence of malignancy considered by Tsushima and Endo<sup>27</sup> (0.67 to 0.64, 0.92 to 0.88, 0.96 to 0.92 and 0.95 to 0.92 for CT, CT then PET, CT then biopsy, and CT then PET then biopsy, respectively). Differences become more marked for the strategies including biopsy at higher prevalences.

f Surgery wedge/resection is 1.0 for wedge resection by exploratory thoracotomy and 0.5 for wedge resection by VATS.

The studies differ more noticeably in assumptions regarding growth of benign nodules. Lejeune *et al.*<sup>28</sup> assumed that none of the benign nodules grows. This does not seem to reflect current understanding of natural history and is likely to underestimate treatment costs for the watch-and-wait strategy. Gambhir *et al.*<sup>22</sup> and Dietlein *et al.*<sup>24</sup> assumed that 10% of benign nodules grow at a constant rate within 2 years of follow-up, undergoing unnecessary surgery. Gould *et al.*<sup>20</sup> assumed a high rate of growth (28%) in the first month of observation, due to infectious or inflammatory processes, with much lower rates (0.5%) subsequently (a cumulative percentage of 36% over 24 months).

All studies that model patient outcomes assumed that those with benign disease have a life expectancy similar to that of the general population, although patients experiencing biopsy or unnecessary surgery are exposed to procedure-related mortality risks. Of the four studies using life expectancies estimated outside the model, two<sup>22,55</sup> were explicit in reporting the approach used to derive life expectancies, using approximations developed by Beck *et al.*<sup>59,60</sup> It was not clear how the other two studies<sup>24,28</sup> derived these values, as they reference several epidemiological references, but also refer to the methods used in the Gambhir *et al.*<sup>22</sup> study. None of these studies explicitly modelled recurrence of disease following resection. As lung cancer deaths due to recurrence should be captured within the epidemiological studies used to estimate life expectancy, these estimates should not be affected by bias. However, it would be likely to underestimate the effect of recurrence on quality of life. Gould *et al.*<sup>20</sup> explicitly modelled disease progression and recurrence in their Markov model, applying stage-specific mortality rates to estimate life expectancy in the model.

### Model input

The inputs for the different economic evaluations are outlined in *Table 6*. The studies vary substantially in the number of diagnostic studies used to derive estimates of test accuracy (ranging from 1 to 24 for PET and from 1 to 12 for CT), and differ substantially in the methods used to synthesise the results of multiple studies. The most common methods used (among those studies for which the method was clearly described) were to base estimated accuracy on a single diagnostic study or simple averaging. Three studies used a single study for CT accuracy,<sup>22,26,27</sup> two used a single study for accuracy of PET,<sup>21,26</sup> and two used simple averaging for PET accuracy.<sup>22,24</sup>

Only Gould *et al.*<sup>20</sup> and Deppen *et al.*<sup>55</sup> conducted formal meta-analyses. Gould *et al.*<sup>20</sup> derived SROC curves (using the Moses–Shapiro–Littenberg method) and selected points on the ROC using the median specificity from studies included in the meta-analysis. Deppen *et al.*<sup>55</sup> used bivariate random-effects regression to determine the diagnostic accuracy of <sup>18</sup>F-FDG-PET/CT (taking diagnostic accuracy of needle biopsy and CT-guided fine-needle aspiration from published literature<sup>61</sup>).

None of the studies reports formal assessment of diagnostic accuracy studies (used to derive estimates of test accuracy used in the models) against prespecified criteria. There is also limited discussion of the appropriateness or quality of the studies providing model inputs, although Keith *et al.*<sup>21</sup> argue in favour of jurisdiction-specific accuracy data. There are striking differences between input values in some studies; for example:

- The CT accuracy used by Gugiatti *et al.*<sup>25</sup> (sensitivity = 0.53, specificity = 0.75) differs substantially from other studies (ranging from 0.965 to 0.999 for sensitivity and 0.53 to 0.65 for specificity).
- The sensitivity of CT follow-up used by Tsushima and Endo<sup>27</sup> (0.56) differs from that of other studies using chest X-ray follow-up (e.g. 1.0<sup>21,26</sup>).
- The procedure-related morbidity and mortality, when reported, varied (e.g. mortality from biopsy ranged from 0.005<sup>20</sup> to 0.2<sup>22,24</sup>).
- The cost inputs vary substantially between studies (see *Table 6* for costs converted to common base of 2016 US\$). CT costs range from US\$144 to US\$597, PET costs range from US\$748 to US\$2693, 'surgery' costs range from US\$7728 to US\$22,311 and DCE-CT costs range from US\$118 to US\$454 (2016 US\$).

### **Sensitivity analysis**

The majority of studies conducted deterministic sensitivity analyses only, including one- and two-way analyses (see *Appendix 9, Table 38*). The exception was Gould *et al.*,<sup>20</sup> who reported conducting a probabilistic analysis with 10,000 model replications, but provided little detail on the inputs included in the analysis or their parameterisation, the way the analysis was conducted or the results of the analysis. Three studies reported very limited sensitivity analyses (in two cases, limited to varying the probability of malignancy and cost of PET<sup>21,26</sup> and, in another, to varying the probability of malignancy only<sup>27</sup>). In studies reporting more extensive sensitivity analyses, it remains unclear how far any of the studies have adequately characterised uncertainty in their models. Most studies appear to use arbitrary ranges for their sensitivity analyses, providing no rationale for the values chosen. Those studies that have included test accuracy in their sensitivity analyses carried out one of the following: conducted separate analyses for sensitivity and specificity,<sup>22,25</sup> simultaneously increased or decreased sensitivity and specificity<sup>22,24</sup> or provided no indication of how they have dealt with the inherent relationship between sensitivity and specificity.<sup>28</sup> Deppen *et al.*<sup>55</sup> conducted a two-way sensitivity analysis for the sensitivity and specificity of <sup>18</sup>F-FDG-PET/CT. It is not clear if this was the only such analysis conducted, or if they included other tests in this analysis.

Gould *et al.*<sup>20</sup> account for correlation between sensitivity and specificity for each test in their probabilistic sensitivity analysis by sampling from logit-normal distribution<sup>62</sup> for specificity and then deriving an accompanying sensitivity from the ROC curve. This suggests a highly deterministic relationship between sensitivity and specificity that is unlikely to represent joint uncertainty. In general, the use of the logit-normal distribution in this analysis reflects a pragmatic decision that would not meet current methodological standards for probabilistic analysis.

### **Economic evidence to support individual tests or diagnostic strategies**

It is difficult to identify estimates of the cost-effectiveness of individual tests because most of the strategies modelled in the studies include combinations of tests (see *Appendix 9, Tables 39 and 40*). Where single tests/interventions have been included (observation,<sup>20,22,24,28</sup> CT only,<sup>20–22,25–27</sup> PET only,<sup>21,24,28,55</sup> biopsy<sup>20,24,55</sup> or immediate surgery<sup>20,22,24,55</sup>), results suggest that watchful waiting may be preferred when nodules are very small or prevalence/prior probability of malignancy is low ( $\leq 10\%$ )<sup>20,22,24</sup> and immediate surgery preferred when prior probability of malignancy is very high ( $> 90\%$ ).<sup>20,22,24</sup> Otherwise, CT is the preferred option.<sup>20,22</sup> The exceptions to this rule are the analyses by Keith *et al.*,<sup>21</sup> Comber *et al.*<sup>26</sup> and Deppen *et al.*<sup>55</sup> However, this is more likely to reflect differences in methodology [modelling an intermediate outcome (test accuracy), choice of comparator (no imaging<sup>21,26</sup> or excluding watchful waiting<sup>55</sup>) and comparing all tests against a common comparator].

Including strategies with sequential testing brings <sup>18</sup>F-FDG-PET/CT into the diagnostic pathway (as follow-up to initial CT). Gould *et al.*<sup>20</sup> report the most sophisticated analysis of such sequences, incorporating both pre-test and post-test probability of malignancy to derive recommendations for diagnostic sequences following initial CT, stratified by diagnostic outcome of CT (see Gould *et al.*,<sup>20</sup> table 2 of the main paper, or see figures 9 and 10 in the appendix for graphical representations of these recommendations).

### **Discussion**

The purpose of this review was twofold:

1. to identify the current state of the economic evidence supporting individual tests or strategies in the characterisation of SPNs; in particular to identify evidence (if any) of the cost-effectiveness of DCE-CT
2. to identify approaches that have been used to estimate the cost-effectiveness of diagnosis and management of SPNs, in particular the methods used to synthesise evidence of diagnostic accuracy and to model longer-term patient outcomes.

In addition to concerns over heterogeneity among comparators and study methods typically identified by reviews of economic evaluations, this review identified a number of methodological flaws in the identification of evidence (lack of systematic searching, pre-determined inclusion/exclusion criteria or quality assessment for model inputs), evidence synthesis (inadequate methods of data pooling) and conduct of economic evaluations (failure to present fully incremental analyses, inadequate specification, conduct or presentation of sensitivity analyses and an absence of properly conducted probabilistic sensitivity analyses).

This review has been able to draw tentative conclusions regarding the cost-effectiveness of individual tests and the relative ordering of test sequences, based on published economic evaluations. However, one important reservation remains when drawing such conclusions. Given that several studies have drawn from a common set of diagnostic studies and adopted broadly similar analytical approaches, false reassurance may result from the similar findings, which may be due to any biases from the similar sources and methods used. In other words, a simple total derived by stating that  $n$  studies found CT to be cost-effective at first screen may be undermined by the fact that the studies are not truly independent. These, and other concerns (such as context dependency and more general issues with generalisability), have led some authors to question the value of systematic reviews of published economic evaluations.<sup>63</sup>

### Relevance to health service decision-making

The perspectives adopted by the majority of the papers included in this review (health service or third-party payer) suggest that the authors of the papers expected their results to be used to aid health-care decision-making. However, the lack of clarity on populations, the range of outcomes presented (cost per case, improved accuracy, life-year gains, and only one study reporting QALY gains), differences in assumed diagnostic pathways (see footnotes to *Table 6*) and the varying jurisdictions represented (USA, France, Germany, Australia, Japan, Italy) limit the relevance of these studies for use in health service decision-making in the UK. The most serious drawback in the analyses and presentation of the results of the studies is the lack of consideration given to the relevance of the comparators used (in terms of the standard of care at the time when the studies were conducted) and the lack of fully incremental analyses. Methodological guidance on the conduct of health technology<sup>22</sup> and diagnostic assessments issued by NICE<sup>40</sup> may be considered reasonable indicators of best practice. These guidance emphasise the importance of using comparators that reflect established clinical practice. The included studies spend little time discussing what might be considered to be the most appropriate comparators for their specific jurisdictions and, in some cases, adopt comparators that they acknowledge do not reflect established clinical practice.<sup>21,26</sup>

The usefulness of these analyses is also limited by the scope of the outcomes adopted. Although QALYs are not a universally adopted outcome measure for technology assessment and reimbursement decisions, the more limited outcome measures adopted in most included studies provided little guidance for strategic decisions between patient or disease groups. As noted earlier, studies reporting cost per unit gain in diagnostic accuracy can only inform decisions regarding the optimal method of characterising SPNs. Such studies also, implicitly, place equal weight on disparate outcomes that may not reflect patient or general public preferences: should equal weight be applied to false-positive test results, implying unnecessary treatment, and false-negative test results, the latter implying a missed treatment opportunity?

### Conclusion

This review of published economic evaluations of diagnostic imaging strategies to characterise SPNs has been able to provide some evidence of the cost-effectiveness of strategies in relation to the probability of malignancy. However, it is not possible to draw firm, quantitative conclusions from the results of the different studies because of differences in assumptions, study methods and perspectives. The principal purpose of this review was to assess the current state of the evidence on the cost-effectiveness of diagnostic strategies for the characterisation of SPNs. In particular,

it intended to identify the current evidence on the role of DCE-CT in the diagnostic pathway and to justify the inclusion of a full economic evaluation as part of the ongoing SPUtNik study. The gaps in the evidence base and the concerns we have highlighted in methodology in a number of published studies suggest that such a study is justified.

## Chapter 6 Main study results

### Centres

A total of 16 centres participated in the SPUtNIK trial: three in Scotland and 13 in England. These 16 sites recruited a total of 380 participants (median per site 21.5 participants, range 2–74 participants). See *Appendix 10* for site details.

### Screened patients

A total of 2541 patients were screened (*Figure 12*). The most common reasons for screen failure were the presence of more than one nodule [ $n = 413$  (19%)], patient declining [ $n = 296$  (14%)], nodule outside the size range [ $n = 306$  (14%)] and malignancy within the previous 2 years [ $n = 264$  (12%)].

### Recruited patients

Of the 2541 patients screened, 380 were recruited to the final study.

### Participant withdrawal

Fifteen participants withdrew consent between recruitment and the performance of one or both imaging techniques (see *Figure 12*). Eight of these participants declined both imaging techniques, and seven withdrew consent prior to DCE-CT. Four participants withdrew consent between the imaging techniques and completion of the 2-year follow-up, prior to receiving a confirmed diagnosis. One participant died prior to receiving a diagnosis; their cause of death was not related to the nodule.

### Participant follow-up

A total of 312 participants completed follow-up with usable information from both imaging modalities. Of these, 205 underwent histological/pathological sampling and 191 out of 312 (61.2%) of the SPNs were lung cancer (see *Table 14*).

### Numbers analysed

A total of 380 participants were recruited to the study; however, six were found to be ineligible because of either multiple nodules ( $n = 4$ ) or having a nodule that was actually outside the size range ( $n = 2$ ). Of the remaining 374 participants, 362 underwent PET/CT and 329 underwent DCE-CT. Of the 12 participants who did not undergo PET/CT, eight declined, two could not undergo PET/CT for technical reasons, one could not undergo PET/CT for a medical reason and one had a nodule that could no longer be detected. Of the 45 participants who did not undergo DCE-CT, 15 declined, 13 had a nodule that could no longer be detected, nine could not undergo DCE-CT for technical reasons and eight could not undergo DCE-CT for medical reasons (see *Figure 12*).

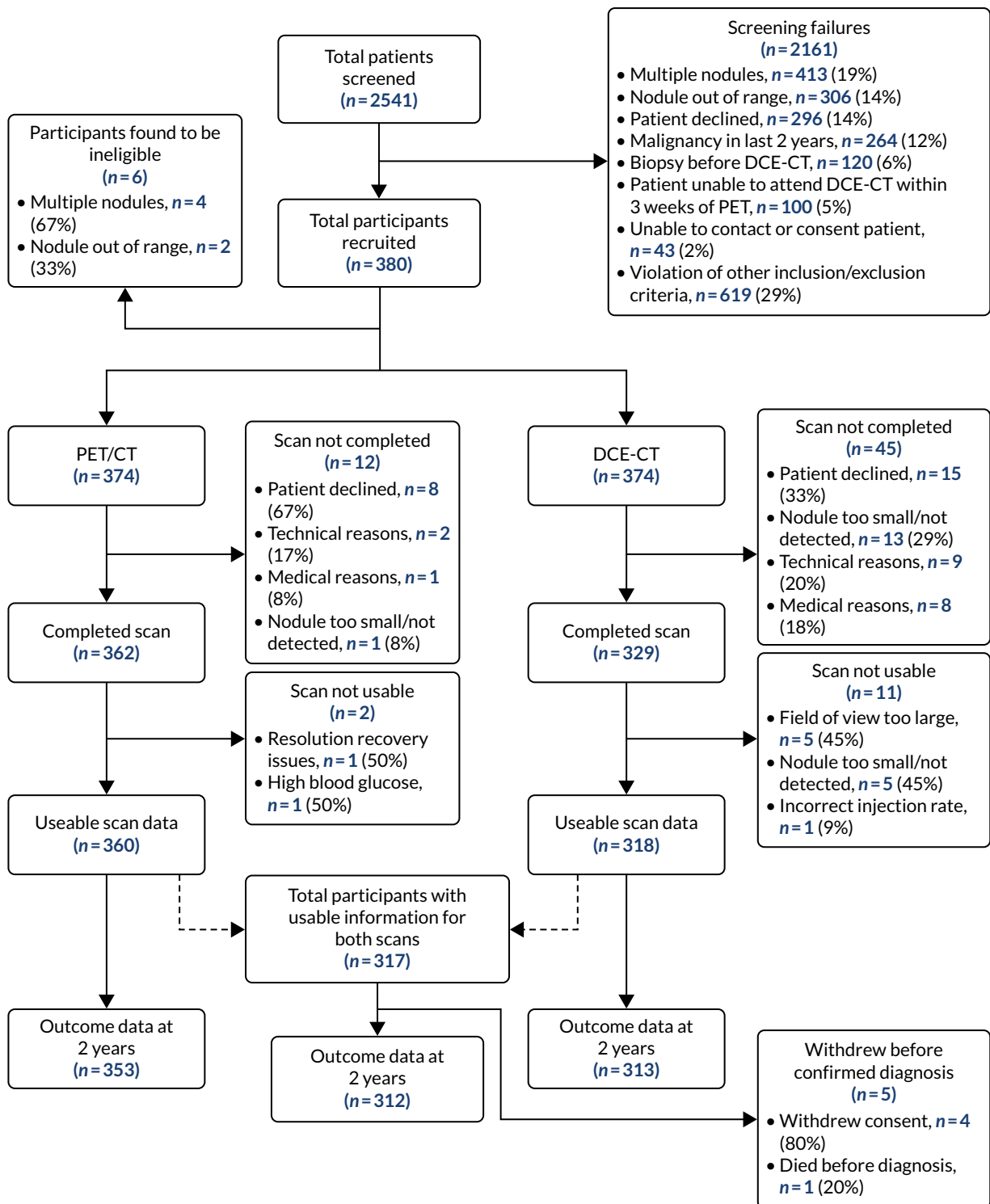


FIGURE 12 The SPUtNIK study standards for the reporting of diagnostic accuracy studies (STARD) flow diagram.

Of the completed scans, two of the PET/CT images were later found to be unusable because of a resolution recovery issue and a higher-than-allowed blood glucose level. Eleven of the completed DCE-CT scans were later found to be unusable, with the field of view being too large in five of the scans, the nodule being too small to report in five of the scans and a single case of an incorrect contrast injection rate. This resulted in 360 participants having usable PET/CT scan data and 318 having usable DCE-CT scan data. A total of 317 participants had usable scan data for both modalities; 312 of these had a complete nodule status at 2 years.

## Baseline data and demographics

Baseline characteristics for all recruited individuals, eligible participants consented to the study and the subset of participants with usable data from both scans and a 2-year outcome status are summarised in *Table 7*. This includes the approximate location of the identified SPN and a targeted medical history.

TABLE 7 Baseline characteristics and medical history of the study participants

Variable (unit)	Main analysis set <sup>a</sup> (N = 312)	Total eligible participants (N = 374)	Total recruited (N = 380)
<b>Sex, n (%)</b>			
Male	165 (53)	199 (53)	201 (53)
Female	147 (47)	175 (47)	179 (47)
<b>Age (years)</b>			
Mean (SD)	68.1 (8.95)	67.9 (8.97)	67.9 (8.92)
Median	69	69	68.5
LQ to UQ	62 to 74	62 to 74	62 to 74
Minimum, maximum	35, 89	35, 89	35, 89
<b>Smoking status, n (%)</b>			
Never-smoker	57 (19)	66 (18)	66 (18)
Ex-smoker	170 (56)	204 (57)	207 (57)
Current smoker	77 (25)	90 (25)	92 (25)
Missing (n)	8	14	15
<b>Location of SPN, n (%)</b>			
Left lower lobe	51 (16)	63 (17)	63 (17)
Left upper lobe	78 (25)	95 (26)	98 (26)
Right lower lobe	73 (23)	77 (21)	77 (21)
Right middle lobe	21 (7)	25 (7)	25 (7)
Right upper lobe	89 (29)	107 (29)	108 (29)
Missing (n)	0	7	9
<b>WHO performance status grade, n (%)</b>			
0: Fully active, able to carry on all pre-disease performance without restriction	151 (49)	179 (48)	183 (49)
1: Restricted in physically strenuous activity, but ambulatory and able to carry out work of a light or sedentary nature, for example light house work, office work	133 (43)	159 (43)	161 (43)
2: Ambulatory and capable of all self-care, but unable to carry out any work activities; up and about for > 50% of waking hours	22 (7)	26 (7)	26 (7)
3: Capable of only limited self-care; confined to bed or chair for > 50% of waking hours	5 (2)	7 (2)	7 (2)
Missing (n)	1	3	3

continued

TABLE 7 Baseline characteristics and medical history of the study participants (continued)

Variable (unit)	Main analysis set <sup>a</sup> (N = 312)	Total eligible participants (N = 374)	Total recruited (N = 380)
<b>Medical history of cardiovascular disease</b>			
Any cardiovascular disease, n (%)	70 (23)	85 (24)	85 (23)
Ischaemic heart disease, n (%)	51 (17)	62 (17)	62 (17)
Valve disease, n (%)	11 (4)	12 (3)	12 (3)
Cardiomyopathy, n (%)	2 (1)	3 (1)	3 (1)
Missing (n)	11	15	16
<b>Medical history of respiratory disease</b>			
Any respiratory disease, n (%)	126 (41)	156 (43)	160 (43)
COPD, n (%)	90 (29)	114 (31)	117 (31)
Asthma, n (%)	29 (9)	37 (10)	38 (10)
Pulmonary fibrosis, n (%)	6 (3)	8 (2)	8 (2)
Other ILD, n (%)	0 (0)	0 (0)	0 (0)
Other, n (%)	17 (6)	21 (6)	21 (6)
Missing (n)	5	8	8
<b>Medical history of inflammatory disease</b>			
Any inflammatory disease, n (%)	65 (21)	78 (21)	79 (21)
Rheumatoid, n (%)	20 (6)	24 (7)	24 (6)
Granulomatosis with polyangiitis, n (%)	1 (0)	2 (1)	2 (1)
Missing (n)	4	6	6
<b>Medical history of infectious disease</b>			
Any infectious disease, n (%)	112 (37)	127 (35)	129 (35)
Histoplasmosis, n (%)	1 (0)	1 (0)	1 (0)
Chickenpox, n (%)	108 (35)	122 (34)	123 (33)
Tuberculosis, n (%)	9 (3)	9 (2)	10 (3)
Missing (n)	6	12	12
<b>Previous exposures</b>			
Any previous exposure, n (%)	63 (21)	76 (21)	77 (21)
Asbestos, n (%)	55 (18)	68 (19)	69 (19)
Coal, n (%)	14 (5)	14 (4)	14 (4)
Silica, n (%)	4 (1)	4 (1)	4 (1)
Missing (n)	14	16	16
<b>Prior malignancy</b>			
Any prior malignancy, n (%)	38 (12)	51 (14)	52 (14)
Missing (n)	6	9	9
COPD, chronic obstructive pulmonary disease; ILD, interstitial lung disease; LQ, lower quartile; SD, standard deviation; UQ, upper quartile; WHO, World Health Organization.			
a Participant has usable PET/CT and DCE-CT scans and has a 2-year outcome status.			

The sex balance of the recruited sample was relatively even, with slightly more male participants (53%). The median age was 69 years, with a range of 35–89 years. Of the recruited participants, 299 out of 380 reported a history of smoking, although only 92 (25%) reported themselves as current smokers (77/312 in the main analysis set). Sixty-six out of 380 (57/312 in the main analysis set) participants were never smokers. The location of the identified SPN was relatively evenly spread across the sites, with the right middle lobe occurring less frequently (7%), and the right upper lobe most frequently (29%). The comorbidities reported were thought to be representative for this age group, suggesting that a representative sample was recruited. Fifty-two out of 380 (38/312 in the main analysis set) participants reported having had a previous malignancy > 2 years before entering the study. The 264 individuals screened for the study with a malignancy within the previous 2 years (see *Figure 12*) were not recruited to the study.

### Baseline computerised tomography, fluorodeoxyglucose positron emission tomography-computerised tomography and dynamic contrast-enhanced computerised tomography results

The timings between the imaging procedures and the imaging results are presented in *Table 8*. Of the 312 participants who successfully completed the 2-year follow-up with usable information for both PET/CT and DCE-CT, 282 (90%) underwent DCE-CT within 2 weeks of PET/CT (median delay 1 day, range 0–32 days). A total of 307 (98%) participants underwent both PET/CT and DCE-CT within 3 weeks of each other. In 154 instances (49%), both the PET/CT and DCE-CT occurred on the same day.

TABLE 8 The CT, DCE-CT and PET/CT information

Scan	Variable (units)	Main analysis set <sup>a</sup> (N = 312)
Timings	Time between DCE-CT and PET/CT (days) <sup>b</sup>	
	Mean (SD)	4.5 (6.31)
	Median	1
	LQ to UQ	0 to 8
	Minimum, maximum	0, 32
	On the same day, n (%)	154 (49)
	Within 1 week ( $\pm 7$ days), n (%)	231 (74)
	Within 2 weeks ( $\pm 14$ days), n (%)	282 (90)
	Within 3 weeks ( $\pm 21$ days), n (%)	307 (98)
	Within 4 weeks ( $\pm 28$ days), n (%)	310 (99)
Missing (n)	0	
Baseline CT	Grade of SPN, n (%)	
	0: Round, well-defined lesion with laminated or popcorn calcification	5 (2)
	1: Inflammatory features	10 (3)
	2: Smooth, well-defined margins, uniform density	59 (20)
	3: Lobulated, spiculated or irregular margins	212 (74)
	4: Evidence of distant metastases	2 (1)
Missing (n)	24	

continued

TABLE 8 The CT, DCE-CT and PET/CT information (continued)

Scan	Variable (units)	Main analysis set <sup>a</sup> (N = 312)
	Evidence of consolidation or inflammation, n (%)	
	No	279 (90)
	Yes	31 (10)
	Missing (n)	2
	Lymph nodes affected, n (%)	
	No	288 (92)
	Yes	24 (8)
	Missing (n)	0
	Evidence of metastases, n (%)	
	No	309 (100)
	Yes	0 (0)
	Missing (n)	3
Follow-up CT	First follow-up CT (N = 150), <sup>c</sup> n (%)	
	Conducted	135 (90)
	Not conducted	15 (10)
	Second follow-up CT (N = 101), <sup>c</sup> n (%)	
	Conducted	85 (84)
	Not conducted	16 (16)
	Third follow-up CT (N = 44), <sup>c</sup> n (%)	
	Conducted	32 (73)
	Not conducted	12 (27)
PET/CT	SUVmax	
	Mean (SD)	4.91 (5.65)
	Median	2.92
	LQ to UQ	1.5 to 6.2
	Minimum, maximum	0, 56.5
	Missing (n)	2
	SUVmean	
	Mean (SD)	2.51 (3.23)
	Median	1.3
	LQ to UQ	0.6 to 3.05
	Minimum, maximum	0, 20.8
	Missing (n)	12
	Grade of SPN on CT, n (%)	
	0: Round, well-defined lesion with laminated or popcorn calcification	6 (2)
	1: Inflammatory features	13 (4)
	2: Smooth, well-defined margins, uniform density	66 (22)

TABLE 8 The CT, DCE-CT and PET/CT information (continued)

Scan	Variable (units)	Main analysis set <sup>a</sup> (N = 312)
	3: Lobulated, spiculated or irregular margins	208 (71)
	4: Evidence of distant metastases	2 (1)
	Missing (n)	17
	Grade of SPN on PET, n (%)	
	0: No visible uptake	52 (17)
	1: Uptake less than mediastinal blood pool	67 (21)
	2: Uptake comparable to mediastinal blood pool	30 (10)
	3: Uptake greater than mediastinal blood pool	161 (52)
	4: Evidence of distant metastases	2 (1)
	Missing (n)	0
	Radiologist's diagnosis of SPN, n (%)	
	Cancer	90 (29)
	Indeterminate	191 (61)
	Non-cancer	31 (10)
	Diagnosis of SPN according to protocol, n (%)	
	Cancer	161 (52)
	Non-cancer	151 (48)
	Lymph nodes affected, n (%)	
	No	269 (87)
	Yes	40 (13)
	Missing (n)	3
	Evidence of metastases, n (%)	
	No	306 (99)
	Yes	4 (1)
	Missing (n)	2
DCE-CT	Peak HU	
	Mean (SD)	130.8 (52.82)
	Median	128
	LQ to UQ	104 to 152
	Minimum, maximum	-16, 578
	Maximum, mean HU	
	Mean (SD)	64.9 (43.72)
	Median	67
	LQ to UQ	43 to 88.9
	Minimum, maximum	-213, 361

continued

TABLE 8 The CT, DCE-CT and PET/CT information (continued)

Scan	Variable (units)	Main analysis set <sup>a</sup> (N = 312)
	Peak enhancement	
	Mean (SD)	48.6 (28.3)
	Median	46.5
	LQ to UQ	29 to 64.5
	Minimum, maximum	0, 179
	Grade of SPN, n (%)	
	0: Round, well-defined lesion with laminated or popcorn calcification	3 (1)
	1: Inflammatory features	12 (4)
	2: Smooth, well-defined margins, uniform density	63 (21)
	3: Lobulated, spiculated or irregular margins	223 (74)
	4: Evidence of distant metastases	0 (0)
	Missing (n)	10
	Radiologist's diagnosis of SPN, n (%)	
	Cancer	51 (17)
	Indeterminate	227 (73)
	Non-cancer	31 (10)
	Missing (n)	3
	Diagnosis according to peak enhancement of $\geq 15$ HU, n (%)	
	Cancer	281 (90)
	Non-cancer	31 (10)
	Diagnosis according to peak enhancement of $\geq 20$ HU, n (%)	
	Cancer	267 (86)
	Non-cancer	45 (14)

LQ, lower quartile; SD, standard deviation; UQ, upper quartile.

a Participant has usable PET/CT and DCE-CT scans, and has a 2-year outcome status.

b Calculated as DCE-CT minus PET/CT.

c Calculated as the absolute value of the difference.

From the baseline CT, most [212 (74%)] of the nodules were classified as 'grade 3: lobulated, spiculated or irregular margins', with evidence of consolidation or inflammation in 31 (10%) of the nodules. Lymph node involvement was present in 24 (8%) of the cases and none of the images showed evidence of distant metastases. Presence of metastases on this baseline CT was one of the study exclusion criteria.

Despite being a low-dose examination, the CT grading from the PET/CT was very similar to that of the baseline CT, with the most common grade remaining grade 3 [208 (71%)]. On grading the PET portion of the examination, 161 (52%) had uptake greater than the mediastinal blood pool (grade 3), 17% had no uptake and 10% had a similar uptake to the mediastinal blood pool (grade 2).

The mean of the SUVmax was 4.75 (with a range of 0–35.3) and the mean of the SUVmean was 2.51 (with a range of 0–20.8). There were slightly more missing data for the SUVmean, as not all sites report this as standard. In only 121 cases (39%) was the radiologist able to make a definitive diagnosis from the PET/CT images, with the remainder (61%) classified as indeterminate. When following the study protocol for determining the status of the PET/CT result, a diagnosis of cancer was reached for 161 (52%) cases. It was not possible to reach an indeterminate result using the rules in our protocol. In 40 (13%) cases, there was lymph node involvement, and in four (1%) there was evidence of metastases.

From the DCE-CT images, the mean of the peak enhancement across all slices was 48.6 HU (range 0–179 HU). The most common grade remained grade 3: lobulated, spiculated or irregular margins [223 (74%)]. For 82 cases (27%), the radiologist involved was able to make a definitive diagnosis from the DCE-CT images; an indeterminate diagnosis was given for the remaining 227 cases (73%). When using a peak enhancement threshold of  $\geq 20$  HU, a diagnosis of cancer was reached for 267 (86%) cases.

The sizes of the SPNs, as measured from the baseline CT, PET/CT and DCE-CT, are presented in *Table 9*. Furthermore, *Table 9* is subdivided by the 2-year outcome status of the SPN (cancer or non-cancer). The mean of the maximum diameter from the baseline CT scan was 15.9 mm, with a range of 8–30 mm. Generally, the sizes from the PET/CT scan are similar to those from the baseline CT scan, with the DCE-CT measurements tending to be smaller. When splitting the results by outcome status, malignant nodules are larger, on average, than the non-cancerous nodules, although there is considerable overlap between the groups.

For a diagnosis of malignancy, histological confirmation was obtained or, if biopsy/resection was not possible, an increase in nodule size with MDT certainty of malignancy.

TABLE 9 The SPN sizes according to each imaging technique and split by cancer status ( $N = 312$ )

SPN	Variable (unit)	Baseline CT	PET/CT	DCE-CT
All nodules ( $N = 312$ )	Transverse diameter (mm)			
	Mean (SD)	15.5 (5.30)	15.6 (6.05)	14.5 (5.97)
	Median	14	14	13
	LQ to UQ	11 to 19	11 to 19	10 to 18
	Minimum, maximum	7, 30	1.2, 32	1.5, 31
	Missing ( $n$ )	6	17	0
	Perpendicular diameter (mm)			
	Mean (SD)	13.4 (5.00)	13.2 (5.29)	12.2 (5.12)
	Median	12	12	11
	LQ to UQ	10 to 16	9 to 16	8 to 15
	Minimum, maximum	4, 30	1.4, 34	1.1, 28
	Missing ( $n$ )	75	33	1
	Maximum diameter (mm)			
	Mean (SD)	15.9 (5.30)	16.0 (6.01)	15.1 (5.84)
	Median	15	15	14
	LQ to UQ	12 to 20	12 to 19	11 to 19
	Minimum, maximum	8, 30	1.4, 34	1.5, 31
	Missing ( $n$ )	6	17	0

continued

TABLE 9 The SPN sizes according to each imaging technique and split by cancer status (N = 312) (continued)

SPN	Variable (unit)	Baseline CT	PET/CT	DCE-CT	
Non-cancer at 2 years (N = 121)	Transverse diameter (mm)				
	Mean (SD)	14.0 (4.31)	13.5 (4.93)	12.3 (4.72)	
	Median	13	12.1	11	
	LQ to UQ	11 to 16	10 to 16	9 to 15	
	Minimum, maximum	8, 30	2, 30	1.5, 30	
	Missing (n)	3	8	0	
	Perpendicular diameter (mm)				
	Mean (SD)	12.2 (4.35)	11.1 (4.02)	10.0 (4.05)	
	Median	11	10	9	
	LQ to UQ	9 to 14	8 to 14	7 to 12.3	
	Minimum, maximum	4, 30	4, 24	1.1, 23	
	Missing (n)	35	16	1	
	Maximum diameter (mm)				
	Mean (SD)	14.3 (4.42)	13.9 (4.92)	12.8 (4.71)	
	Median	13	13	12	
	LQ to UQ	11 to 16	10 to 16	9 to 16	
	Minimum, maximum	8, 30	2, 30	1.5, 30	
	Missing (n)	3	8	0	
	Cancer at two years (N = 191)	Transverse diameter (mm)			
		Mean (SD)	16.4 (5.65)	16.8 (6.32)	15.8 (6.27)
Median		15.5	15	14	
LQ to UQ		12 to 20	12 to 21	11 to 20	
Minimum, maximum		7, 29	1.2, 32	4, 31	
Missing (n)		3	9	0	
Perpendicular diameter (mm)					
Mean (SD)		14.1 (5.23)	14.4 (5.58)	13.5 (5.25)	
Median		13	14	12	
LQ to UQ		10 to 18	10 to 18	10 to 17	
Minimum, maximum		6, 27	1.4, 34	4, 28	
Missing (n)		40	17	0	
Maximum diameter (mm)					
Mean (SD)		16.9 (5.57)	17.4 (6.24)	16.6 (6.03)	
Median		16	16.5	15	
LQ to UQ		12 to 21	13 to 21.6	12 to 21	
Minimum, maximum		8, 29	1.4, 34	4, 31	
Missing (n)		3	9	0	

LQ, lower quartile; SD, standard deviation; UQ, upper quartile.

## Prevalence of cancer

At the end of 2 years' follow-up, 191 (61%) participants had confirmation of lung cancer. *Table 10* shows that there is a variation in the prevalence across the recruiting sites, with values ranging from 11 to 80%, although only three sites have a prevalence of < 42%. There was also a difference in the total number of participants recruited across these sites, meaning that some of these estimated percentages will be imprecise. Cancers were confirmed by histological diagnosis, except for participants undergoing SABR, or when biopsy was not considered possible.

*Table 11* shows the type of cancers and non-malignant nodules by smoking status. The most common type of cancer was non-small-cell (76%), with adenocarcinoma making up 74% of that subgroup. The squamous cell cancers (30/191) were all found in the current or ex-smokers groups. Benign disease was confirmed by biopsy in 27 cases and by up to 2 years' follow-up with CT for 94 patients. The single SPN in the 'other' category was part of a diagnosis of diffuse neuroendocrine cell hyperplasia.

A total of 178 participants had findings on their PET/CT other than their pulmonary nodule. Of these findings, the majority were pulmonary findings that would have been detected on the initial recruitment CT, such as emphysema, fibrosis or pleural plaques, and were not considered to be significant incidental findings. Seventy-three findings were extrapulmonary; the locations of these are summarised in *Table 12*. In the head and neck group, 11 participants had abnormal FDG uptake in the thyroid, five had abnormal uptake in a parotid, six had abnormal uptake in the nasopharynx/tonsils and the remaining three had abnormal uptake in the sinuses or soft tissues. Seven indeterminate breast lesions and seven hiatus hernias were detected, and nine with gastro-oesophageal uptake suggestive of gastro-oesophageal reflux disease were detected.

TABLE 10 Confirmation of malignant SPN following 2 years' follow-up (N = 312)

Site	n/N (%)
Confirmation of malignant SPN	191/312 (61)
Royal Papworth Hospital	45/71 (63)
Leeds	30/42 (71)
Glasgow	29/40 (73)
Aberdeen	24/34 (71)
Brighton	8/19 (42)
Southampton	9/19 (47)
UCLH	9/18 (50)
Oxford	10/19 (53)
Worcester	5/9 (56)
Manchester	8/10 (80)
Nottingham	11/15 (67)
Leicester	1/9 (11)
Hastings	1/3 (33)
Weston	0/2 (0)
Edinburgh	1/2 (50)
UCLH, University College London Hospitals NHS Foundation Trust.	

TABLE 11 Final nodule diagnosis by smoking status (N = 312)

Two-year malignancy status	Classification of nodule	Smoking status, n (%)			Total
		Never smokers	Ex-smokers	Current smokers	
Malignant	Number of participants <sup>a</sup>	27	102	57	191
	Non-small-cell lung cancer	15 (56)	78 (76)	47 (82)	145 (76)
	Adenocarcinoma	14 (93)	55 (71)	34 (72)	107 (74)
	Squamous cell carcinoma	0 (0)	20 (26)	10 (21)	30 (21)
	Large-cell undifferentiated	1 (7)	0 (0)	1 (2)	2 (1)
	Not otherwise specified	0 (0)	3 (4)	2 (4)	6 (4)
	Carcinoid tumour	8 (30)	4 (4)	0 (0)	12 (6)
	Small-cell lung cancer	0 (0)	6 (6)	1 (2)	7 (4)
	SABR	0 (0)	6 (6)	5 (9)	11 (6)
	Radiological diagnosis only	0 (0)	7 (7)	2 (4)	9 (5)
	Other	4 (15)	1 (1)	1 (2)	6 (3)
	No further information provided	0 (0)	0 (0)	1 (2)	1 (1)
	Benign	Number of participants <sup>b</sup>	30	68	20
Benign nodule/lesion		17 (57)	36 (53)	7 (35)	61 (50)
Hamartoma		7 (23)	13 (19)	5 (25)	26 (21)
Infection/inflammation		3 (10)	15 (22)	5 (25)	24 (20)
Other		1 (3)	0 (0)	0 (0)	1 (1)
No further information provided		2 (7)	4 (6)	3 (15)	9 (7)

a Five participants were missing a smoking status.

b Three participants were missing a smoking status.

TABLE 12 Extrathoracic findings from PET/CT (N = 312)

Location	Total, n (%)
Head and neck	25 (8)
Breast	7 (2)
Oesophagus	9 (3)
Hernia	7 (2)
Colon	25 (8)
Kidney	4 (1)
Vascular	8 (2)
Other	5 (2)

The 25 colonic findings were all indeterminate focal hotspots within the colon and rectum, with the sigmoid colon most commonly affected. In two participants, the increased uptake was multifocal, involving two sites. Three indeterminate kidney lesions were identified, with a single obstructive ureteric calculus. Six thoracic/abdominal aortic aneurysms were detected, requiring follow-up. Further incidental findings included two participants with diffuse large- and medium-vessel uptake suggestive of a vasculitis (one with polyarticular uptake consistent with an inflammatory arthritis and the other with skin plaque uptake consistent with psoriasis). It was not possible to follow up all of these abnormalities to determine if these were confirmed pathology or incidental findings, so no further analysis was performed.

## Diagnostic accuracy of positron emission tomography-computerised tomography and dynamic contrast-enhanced computerised tomography

The various gradings from both the PET/CT and DCE-CT are cross-tabulated against the outcome status of the nodule at 2 years in *Table 13*. The outcome status for the non-cancers has been further subdivided into those who received a negative biopsy and those who did not have a biopsy. This shows the performance of the various diagnosis gradings. *Table 14* displays the diagnostic accuracy characteristics of the key diagnosis gradings at the pre-defined cut-off points. It displays sensitivity, specificity, NPV, PPV and ODA, with 95% CIs.

When comparisons were made according to nodule sizes of 8–15 mm, just over 15 mm to 20 mm, and just over 20 mm to 30 mm, the sensitivity of PET/CT increased with increasing nodule size,

TABLE 13 Grading of SPN and nodule growth by final outcome status (N = 312)

Imaging technique	Grade of SPN	Cancer (N = 191), n (%)	Non-cancer (no biopsy) (N = 94), n (%)	Non-lung cancer (biopsy) (N = 27), n (%)
Staging CT	0: Round, well-defined lesion with laminated or popcorn calcification	4 (80)	1 (20)	0 (0)
	1: Inflammatory features	5 (50)	5 (50)	0 (0)
	2: Smooth, well-defined margins, uniform density	16 (27)	34 (58)	9 (15)
	3: Lobulated, spiculated or irregular margins	149 (70)	45 (21)	18 (8)
	4: Evidence of distant metastases	2 (100)	0 (0)	0 (0)
	Missing	15 (63)	9 (38)	0 (0)
PET/CT (PET grading)	0: No visible uptake	6 (12)	44 (85)	2 (4)
	1: Uptake less than mediastinal blood pool	26 (39)	28 (42)	13 (19)
	2: Uptake comparable to mediastinal blood pool	15 (50)	11 (37)	4 (13)
	3: Uptake greater than mediastinal blood pool	142 (88)	11 (7)	8 (5)
	4: Evidence of distant metastases	2 (100)	0 (0)	0 (0)

continued

TABLE 13 Grading of SPN and nodule growth by final outcome status (N = 312) (continued)

Imaging technique	Grade of SPN	Cancer (N = 191), n (%)	Non-cancer (no biopsy) (N = 94), n (%)	Non-lung cancer (biopsy) (N = 27), n (%)
PET/CT (CT grading)	0: Round, well-defined lesion with laminated or popcorn calcification	0 (0)	6 (100)	0 (0)
	1: Inflammatory features	8 (62)	5 (38)	0 (0)
	2: Smooth, well-defined margins, uniform density	22 (33)	34 (52)	10 (15)
	3: Lobulated, spiculated or irregular margins	154 (74)	39 (19)	15 (7)
	4: Evidence of distant metastases	2 (100)	0 (0)	0 (0)
	Missing	5 (29)	10 (59)	2 (12)
PET/CT (radiologist's diagnosis)	Cancer	82 (91)	3 (3)	5 (6)
	Indeterminate	105 (55)	67 (35)	19 (10)
	Non-cancer	4 (13)	24 (77)	3 (10)
PET/CT (protocol)	Cancer	139 (86)	13 (8)	9 (6)
	Non-cancer	52 (34)	81 (54)	18 (12)
DCE-CT	Maximum mean enhancement of < 15 HU	6 (19)	21 (68)	4 (13)
	Maximum mean enhancement of ≥ 15 HU	185 (66)	73 (26)	23 (8)
	Maximum mean enhancement < 20 HU	9 (20)	32 (71)	4 (9)
	Maximum mean enhancement ≥ 20 HU	182 (68)	62 (23)	23 (9)
DCE-CT (CT grading)	0: Round, well defined lesion with laminated or popcorn calcification	1 (33)	2 (67)	0 (0)
	1: Inflammatory features	5 (42)	7 (58)	0 (0)
	2: Smooth, well-defined margins, uniform density	23 (36)	35 (55)	6 (9)
	3: Lobulated, spiculated or irregular margins	159 (71)	44 (20)	20 (9)
	4: Evidence of distant metastases	0 (0)	0 (0)	0 (0)
	Missing	3 (30)	6 (60)	1 (10)
DCE-CT (radiologist's diagnosis)	Cancer	45 (88)	5 (10)	1 (2)
	Indeterminate	138 (61)	66 (29)	23 (10)
	Non-cancer	5 (16)	23 (74)	3 (10)
	Missing	3 (100)	0 (0)	0 (0)
Difference in nodule size (staging CT to any follow-up CT)	Evidence of nodule growth, n/N (%)	21/46 (46)	21/77 (27)	6/14 (43)

with a corresponding drop in specificity, and an increase in diagnostic accuracy from 75.5% to 83.1% (see Appendix 11) (Figures 13 and 14). However, when doing the same analysis with DCE-CT, there was relatively little change in the sensitivity or specificity with increasing nodule size, although there was a slight improvement in diagnostic accuracy.

The following study objectives are addressed in Table 14:

- Primary objective 1: to determine, with high precision, the diagnostic performances of DCE-CT and PET/CT in the NHS for the characterisation of SPNs.
- Secondary objective 1: to assess, in an NHS setting, the incremental value of incorporating the CT appearances of a SPN into the interpretation of integrated PET/CT examinations.
- Secondary objective 2 (part 1): to assess whether or not combining DCE-CT with PET/CT is more accurate in the characterisation of SPNs than either test used alone or in series.

The imaging technique with the highest sensitivity is DCE-CT (see Table 14), which achieves a value of 96.9% when using an enhancement cut-off point of  $\geq 15$  HU, and 95.3% when using an enhancement of  $\geq 20$  HU, whereas PET/CT achieved a sensitivity of only 72.8%. However, this increased sensitivity is associated with low specificity, whereby DCE-CT achieves only 20.7% and 29.8% for enhancement cut-off points of  $\geq 15$  HU and  $\geq 20$  HU, respectively, compared with 81.8% for PET/CT.

When assessing the incremental benefit of adding the CT results on top of the PET/CT results (see Table 14), both the PET/CT values based on the PET grading alone and the SUVmax of  $\geq 2.5$  perform

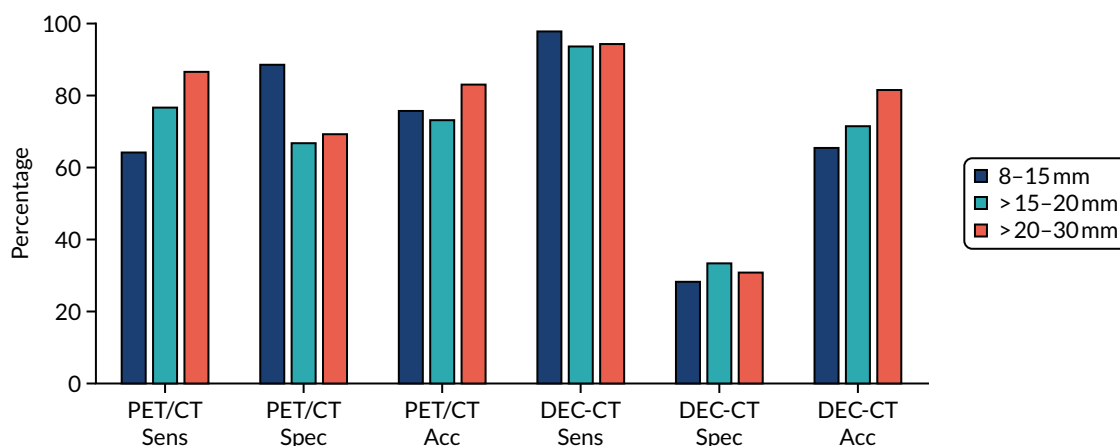


FIGURE 13 Sensitivity, specificity and diagnostic accuracy by nodule size: PET/CT and DCE-CT. Acc, accuracy; Sens, sensitivity; Spec, specificity.

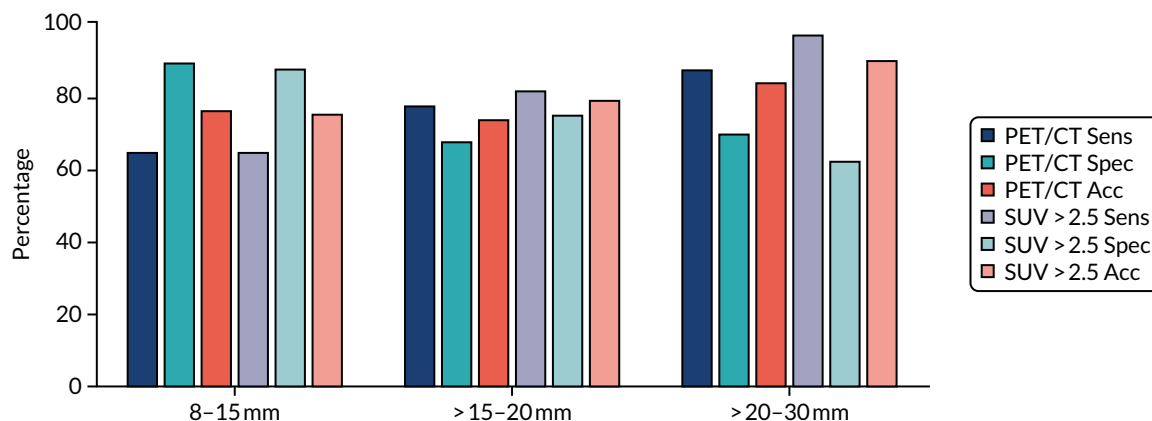


FIGURE 14 Sensitivity, specificity and diagnostic accuracy by nodule size: PET/CT and SUVmax > 2.5. Acc, accuracy; Sens, sensitivity; Spec, specificity.

TABLE 14 Diagnostic performance of each imaging technique: predefined thresholds (N = 312)

Objective	Imaging technique	Sensitivity, n/N (%) (95% CI)	Specificity, n/N (%) (95% CI)	NPV, n/N (%) (95% CI)	PPV, n/N (%) (95% CI)	ODA, n/N (%) (95% CI)
Primary objective 1	DCE-CT (maximum enhancement of $\geq 15$ HU)	185/191 (96.9) (93.3% to 98.6%)	25/121 (20.7) (14.4% to 28.7%)	25/31 (80.7) (63.7% to 90.8%)	185/281 (65.8) (60.1% to 71.1%)	210/312 (66.3) (61.9% to 72.3%)
	DCE-CT (maximum enhancement of $\geq 20$ HU)	182/191 (95.3) (91.3% to 97.5%)	36/121 (29.8) (22.3% to 38.4%)	36/45 (80.0) (66.2% to 89.1%)	182/267 (68.2) (62.4% to 73.5%)	218/312 (69.9) (64.6% to 74.7%)
	PET/CT (based on PET and CT grading)	139/191 (72.8) (66.1% to 78.6%)	99/121 (81.8) (74.0% to 87.7%)	99/151 (65.6) (57.7% to 72.7%)	139/161 (86.3) (80.2% to 90.8%)	238/312 (76.3) (71.3% to 80.7%)
Secondary objective 1	PET/CT (based on PET grading alone)	144/191 (75.4) (68.8% to 81.0%)	102/121 (84.3) (76.8% to 89.7%)	102/149 (68.5) (60.6% to 75.4%)	144/163 (88.3) (82.5% to 92.4%)	246/312 (78.9) (74.0% to 83.0%)
	PET/CT (N = 310) (based on a SUVmax of $\geq 2.5$ )	146/191 (76.4) (69.9% to 81.9%)	97/119 (81.5) (73.6% to 87.5%)	97/142 (68.3) (60.3% to 75.4%)	146/168 (86.9) (81.0% to 91.2%)	243/310 (78.4) (73.5% to 82.6%)
Secondary objective 2	Combination of DCE-CT and PET/CT	134/191 (70.2) (63.3% to 76.2%)	101/121 (83.5) (75.8% to 89.0%)	101/158 (63.9) (56.2% to 71.0%)	134/154 (87.0) (80.8% to 91.4%)	235/312 (75.3) (70.3% to 79.8%)

better than the combined PET/CT results in all areas. This suggests that, in this sample, there is no evidence that adding the CT component is worthwhile.

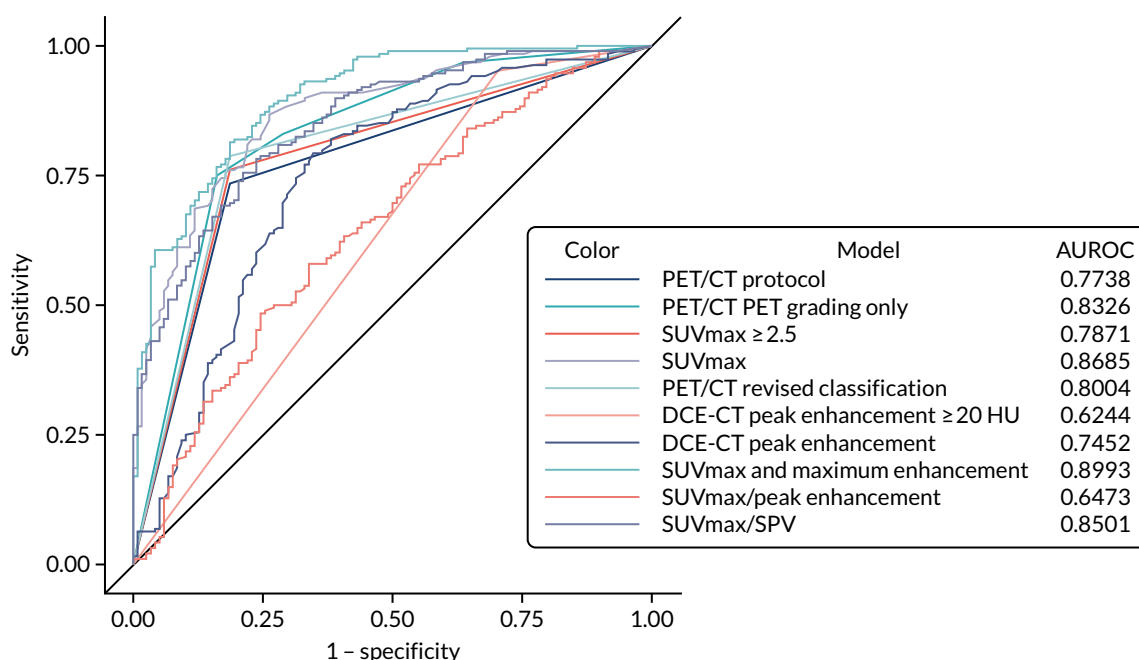
When looking at the combination of DCE-CT with PET/CT (see *Table 14*), the performance is similar to that of PET/CT, but with a slightly lower sensitivity of 70.2% and a higher specificity of 83.5%. This results in the NPV being slightly lower, at 63.9%, and the PPV being marginally higher, at 87.0%. Considering that this combination is equivalent to using one of the imaging techniques and then using the other only if the first was positive (as to be test positive requires a positive result on both modalities), this potentially opens up a cost-saving approach, as DCE-CT is less costly (see *Chapter 7*) than PET/CT. The ODA is lower, at 75.3%, for a combined approach than for PET/CT alone (76.3%), but this small drop in performance may be deemed acceptable.

### Combination of dynamic contrast-enhanced computerised tomography with positron emission tomography-computerised tomography

*Figure 15* shows ROC curves for the prespecified rules and the best-performing combinations following exploratory modelling with logistic regression using the key elements from the imaging scans (including nodule sizes from the scans) to see if a better combination of PET/CT and DCE-CT results is possible. Eight different models are displayed in the ROC curve alongside their AUROCs. Higher AUROCs indicate a better diagnostic performance across the full range of possible cut-off points.

*Table 15* shows a tabulation of the diagnostic performance for the list of models presented in *Figure 15*. It also includes a model comparison (compared with PET/CT) and 95% CIs for both the AUROCs and differences in the AUROCs.

*Figure 15* and *Table 15* show that the best-performing models are the SUVmax for a single variable model and the combination of SUVmax with DCE-CT peak enhancement for a combination of multiple variables. Both of these two diagnostic models produce statistically significantly better AUROCs than



**FIGURE 15** The ROC curves for the diagnostic performance of the prespecified and best-performing exploratory models ( $N = 300$ ).

TABLE 15 Diagnostic performance of the prespecified and best-performing exploratory models (N = 306<sup>a</sup>)

Model	AUROC curve	95% CI	p-value
PET/CT: protocol	0.7738	0.7264 to 0.8212	-
PET/CT: based on PET grade alone	0.8326	0.7872 to 0.8780	-
SUVmax of $\geq 2.5$	0.7871	0.7404 to 0.8338	-
SUVmax	0.8685	0.8281 to 0.9089	-
PET/CT: revised classification	0.8004	0.7545 to 0.8463	-
DCE-CT peak enhancement of $\geq 20$ HU	0.6244	0.5802 to 0.6685	-
DCE-CT peak enhancement	0.7452	0.6856 to 0.8049	-
SUVmax and DCE-CT peak enhancement	0.8993	0.8647 to 0.9340	-
SUVmax/DCE-CT peak enhancement	0.6473	0.5836 to 0.7110	-
SUVmax/SPV	0.8501	0.8076 to 0.8926	-
Difference (PET/CT: based on PET grade alone minus PET/CT: protocol)	0.0588	0.0284 to 0.0892	0.0002
Difference (SUVmax of $\geq 2.5$ minus PET/CT: protocol)	0.0133	-0.0270 to 0.0536	0.5177
Difference (SUVmax minus PET/CT: protocol)	0.0947	0.0586 to 0.1309	< 0.0001
Difference (PET/CT: revised classification minus PET/CT: protocol)	0.0266	0.0105 to 0.0427	0.0012
Difference (DCE-CT peak enhancement of $\geq 20$ HU minus PET/CT: protocol)	-0.1494	-0.2084 to -0.0905	< 0.0001
Difference (DCE-CT peak enhancement minus PET/CT: protocol)	-0.0286	-0.1019 to 0.0448	0.4453
Difference (SUVmax and DCE-CT peak enhancement minus PET/CT: protocol)	0.1255	0.0850 to 0.1661	< 0.0001
Difference (SUVmax/DCE-CT peak enhancement minus PET/CT: protocol)	-0.1265	-0.1898 to -0.0633	< 0.0001
Difference (SUVmax/SPV minus PET/CT: protocol)	0.0763	0.0352 to 0.1173	0.0003

a Subset of participants with complete information on all of the included variables. Shading indicates comparison of AUROCs between the different diagnostic strategies.

the PET/CT protocol that we described earlier (based on the combination of PET/CT and CT elements). For absolute optimum performance, the combination of both SUVmax and DCE-CT peak enhancement performs the best, but the additional performance will be offset by any additional cost and complexity of needing to perform both imaging techniques. However, it does suggest that these scanning technologies can be used in combination to benefit performance.

For ease of comparison, the best cut-off options for all of the models are presented together in *Table 16* (including both a 90% minimum sensitivity cut-off point and a balanced sensitivity and specificity cut-off point). From *Table 16*, it can be seen that changing the SUVmax cut-off point to using a threshold of  $\geq 1.8$  produces an increased performance of 91.0% for sensitivity, 63.0% for specificity, 81.5% for NPV, 79.7% for PPV and 80.3% for ODA, compared with using the 2.5 cut-off point (see *Table 14*). The best-performing single cut-off point involves combining the SUVmax with DCE-CT peak enhancement and using a threshold probability of  $\geq 0.43$ . This produces sensitivity of 90.5%, specificity of 68.9%, a NPV of 82.0%, a PPV of 82.2% and ODA of 82.1%. There is very little to choose between the 90% threshold and optimum cut-off point for this model, with both being very close and performing similarly across the spectrum of values; however, the 90% sensitivity cut-off point correctly classifies one additional participant, compared with the optimum cut-off point (82.1% ODA vs. 81.8% ODA). However, for these models to be used in practice, they would require external validation to show that this was not just an overly optimised performance in the data on which the model was generated.

TABLE 16 Diagnostic performance of the prespecified and best-performing exploratory models (N = 312)

Rule	Imaging technique	Sensitivity, n/N (%) (95% CI)	Specificity, n/N (%) (95% CI)	NPV, n/N (%) (95% CI)	PPV, n/N (%) (95% CI)	ODA, n/N (%) (95% CI)
At least 90% sensitivity (when possible)	PET/CT (based on PET and CT grading)	139/191 (72.8) (66.1% to 78.6%)	99/121 (81.8) (74.0% to 87.7%)	99/151 (65.6) (57.7% to 72.7%)	139/161 (86.3) (80.2% to 90.8%)	238/312 (76.3) (71.3% to 80.7%)
	PET/CT (based on PET grading alone)	185/191 (96.9) (93.3% to 98.6%)	46/121 (38.0) (29.9% to 46.9%)	46/52 (88.5) (77.0% to 94.6%)	185/260 (71.2) (65.4% to 76.3%)	231/312 (74.0) (68.9% to 78.6%)
	PET/CT (N = 310) (based on a SUVmax of $\geq 2.5$ )	146/191 (76.4) (69.9% to 81.9%)	97/119 (81.5) (73.6% to 87.5%)	97/142 (68.3) (60.3% to 75.4%)	146/168 (86.9) (81.0% to 91.2%)	243/310 (78.4) (73.5% to 82.6%)
	SUVmax (N = 309) (positive if $\geq 1.8$ )	173/190 (91.0) (86.1% to 94.3%)	75/119 (63.0) (54.1% to 71.2%)	75/92 (81.5) (72.4% to 88.1%)	173/217 (79.7) (73.9% to 84.5%)	248/309 (80.3) (75.5% to 84.3%)
	PET/CT (revised classification system)	151/191 (79.1) (72.7% to 84.2%)	99/121 (81.8) (74.0% to 87.7%)	99/139 (71.2) (63.2% to 78.1%)	151/173 (87.3) (81.5% to 91.5%)	250/312 (80.1) (75.4% to 84.2%)
	DCE-CT (maximum enhancement $\geq 20$ HU)	182/191 (95.3) (91.3% to 97.5%)	36/121 (29.8) (22.3% to 38.4%)	36/45 (80.0) (66.2% to 89.1%)	182/267 (68.2) (62.4% to 73.5%)	218/312 (69.9) (64.6% to 74.7%)
	DCE-CT peak enhancement (positive if $\geq 25$ HU) (N = 311)	176/190 (92.6) (88.0% to 95.6%)	47/121 (38.8) (30.6% to 47.7%)	47/61 (77.1) (65.1% to 85.8%)	176/250 (70.4) (64.5% to 75.7%)	223/311 (71.7) (66.5% to 76.4%)
	SUVmax and DCE-CT peak enhancement (positive if probability $\geq 0.43$ ) (N = 308)	171/189 (90.5) (85.5% to 93.9%)	82/119 (68.9) (60.1% to 76.5%)	82/100 (82.0) (73.3% to 88.3%)	171/208 (82.2) (76.4% to 86.8%)	253/308 (82.1) (77.5% to 86.0%)
	SUVmax/DCE-CT peak enhancement (positive if probability $\geq 0.599$ ) (N = 310)	172/191 (90.1) (85.0% to 93.5%)	28/119 (23.5) (16.8% to 31.9%)	28/47 (59.6) (45.3% to 72.3%)	172/263 (65.4) (59.5% to 70.9%)	200/310 (64.5) (59.0% to 69.6%)
	SUVmax/SPV (positive if probability $\geq 0.415$ ) (N = 308)	171/190 (90.0) (84.9% to 93.5%)	72/118 (61.0) (52.0% to 69.3%)	72/91 (79.1) (69.7% to 86.2%)	171/217 (78.8) (72.9% to 83.7%)	243/308 (78.9) (74.0% to 83.1%)

continued

TABLE 16 Diagnostic performance of the prespecified and best-performing exploratory models (N = 312) (continued)

Rule	Imaging technique	Sensitivity, n/N (%) (95% CI)	Specificity, n/N (%) (95% CI)	NPV, n/N (%) (95% CI)	PPV, n/N (%) (95% CI)	ODA, n/N (%) (95% CI)
Best balance of sensitivity and specificity	PET/CT (based on PET and CT grading)	139/191 (72.8) (66.1% to 78.6%)	99/121 (81.8) (74.0% to 87.7%)	99/151 (65.6) (57.7% to 72.7%)	139/161 (86.3) (80.2% to 90.8%)	238/312 (76.3) (71.3% to 80.7%)
	PET/CT (based on PET grading alone)	144/191 (75.4) (68.8% to 81.0%)	102/121 (84.3) (76.8% to 89.7%)	102/149 (68.5) (60.6% to 75.4%)	144/163 (88.3) (82.5% to 92.4%)	246/312 (78.9) (74.0% to 83.0%)
	PET/CT (N = 310) (based on a SUVmax of $\geq 2.5$ )	146/191 (76.4) (69.9% to 81.9%)	97/119 (81.5) (73.6% to 87.5%)	97/142 (68.3) (60.3% to 75.4%)	146/168 (86.9) (81.0% to 91.2%)	243/310 (78.4) (73.5% to 82.6%)
	SUVmax (N = 309) (positive if $\geq 2.3$ )	153/190 (80.5) (74.3% to 85.5%)	93/119 (78.2) (69.9% to 84.6%)	93/130 (71.5) (63.3% to 78.6%)	153/179 (85.5) (79.6% to 89.9%)	246/309 (79.6) (74.8% to 83.7%)
	PET/CT (revised classification system)	151/191 (79.1) (72.7% to 84.2%)	99/121 (81.8) (74.0% to 87.7%)	99/139 (71.2) (63.2% to 78.1%)	151/173 (87.3) (81.5% to 91.5%)	250/312 (80.1) (75.4% to 84.2%)
	DCE-CT (maximum enhancement $\geq 20$ HU)	182/191 (95.3) (91.3% to 97.5%)	36/121 (29.8) (22.3% to 38.4%)	36/45 (80.0) (66.2% to 89.1%)	182/267 (68.2) (62.4% to 73.5%)	218/312 (69.9) (64.6% to 74.7%)
	DCE-CT peak enhancement (positive if $\geq 38.5$ ) (N = 311)	147/190 (77.4) (70.9% to 82.7%)	80/121 (66.1) (57.3% to 73.9%)	80/123 (65.0) (56.3% to 72.9%)	147/188 (78.2) (71.8% to 83.5%)	227/311 (73.0) (67.8% to 77.6%)
	SUVmax and DCE-CT peak enhancement (positive if probability $\geq 0.53$ ) (N = 308)	160/189 (84.7) (78.8% to 89.1%)	92/119 (77.3) (69.0% to 83.9%)	92/121 (76.0) (67.7% to 82.8%)	160/187 (85.6) (79.8% to 89.9%)	252/308 (81.8) (77.1% to 85.7%)
	SUVmax/DCE-CT peak enhancement (positive if probability $\geq 0.601$ ) (N = 310)	153/191 (80.1) (73.9% to 85.2%)	44/119 (37.0) (28.8% to 45.9%)	44/82 (53.7) (42.9% to 64.1%)	153/228 (67.1) (60.8% to 72.9%)	197/310 (63.6) (58.1% to 68.7%)
	SUVmax/SPV (positive if probability $\geq 0.542$ ) (N = 308)	141/190 (74.2) (67.6% to 79.9%)	93/118 (78.8) (70.6% to 85.2%)	93/142 (65.5) (57.4% to 72.8%)	141/166 (84.9) (78.7% to 89.6%)	234/308 (76.0) (70.9% to 80.4%)

Of the predefined combination models, the SUVmax/SPV outperforms the SUVmax/DCE-CT peak enhancement quite substantially across the full spectrum. The SUVmax/SPV model performs well and reaches close to the 'best' model, with a sensitivity of 90.0%, a specificity of 61.0%, a NPV of 79.1%, a PPV of 78.8% and ODA of 78.9% at its optimum threshold. All of the values for this model are, however, worse than for the SUVmax and DCE-CT peak enhancement combination that was produced via a logistic regression model.

## Assessing barriers to recruitment and actions taken to increase enrolment

### Overview

The SPUtNik trial opened its first site to recruitment in December 2012. It was expected that seven other sites would open within months and that a total of 375 patients would be recruited within 18 months. However, by January 2014, only six of the expected sites had opened, with a total of 44 patients recruited and only 6 months of recruitment left. There had been extensive problems with site openings and recruitment was far below expected numbers based on the feasibility studies. The recovery plan accepted by the HTA programme was based on opening additional sites in a more realistic time scale with more feasible recruitment targets. The new timelines aimed to recruit 380 participants by October 2016, with complete 2-year follow-up data on 375 patients by December 2018. From a patient perspective, participation in the SPUtNik trial was straightforward and few patients who were approached declined.

The three main barriers to recruitment were as follows:

1. Delays in opening additional sites because of different routes for accessing research support and excess treatment costs at each of the trusts. Research support and excess treatment costs are funded by different bodies: the National Institute for Health Research (NIHR) Clinical Research Network (CRN) and the Clinical Commissioning Group (CCG), respectively. A new trial could not be adopted at a site if annual budgets had already been allocated.
2. Issues arising from trust-to-trust communication and transfer of patient data, which were responsible, in part, for the overestimation of patient recruitment in the initial feasibility studies.
3. Poor communication between chest physicians, radiologists and CRN research support teams initially meant that potential patients were missed or flagged too late for recruitment.

### Clinical Research Network

The CRN appeared to work very differently at some of the SPUtNik sites. An example of good practice was in Southampton, where the CRN worked closely with the local CCG so that excess treatment costs were agreed. In contrast, in Hastings, a separate application to the CCG was needed. In some sites, the CRN provided research nurses to help with the site-specific information, ethics, and research and development (R&D) forms, facilitating rapid opening of a site, whereas, in Oxford, research nurse support was delayed for 1 year.

At one site, CRN research nurses were reluctant to take consent. At another site, no CRN support was allocated, whereas Aberdeen had a research nurse assigned to lead on the trial and cover all aspects, and the Glasgow site was assigned a lead nurse with a team of four or five nurses, which led to smooth running of recruitment and data collection.

There was variation as to what CRN research nurses were expected to do at each site; some were not involved in completing the site-specific information and R&D forms, and others insisted that their role was purely for the recruitment process and refused to collect follow-up data.

### ***Additional financial support at sites***

The CRN at some sites covered clerical costs associated with the trial, such as the costs of printing and posting patient information leaflets, whereas the trial had to pay these costs at another site. Many of the radiology departments were willing to cover the costs of scan anonymisation and to transfer images to SCTU. However, some sites insisted on a payment of £10 per scan for the initial CT scans.

### ***Excess treatment costs/dynamic contrast-enhanced computerised tomography***

Dynamic contrast-enhanced computerised tomography was classed as an excess treatment cost, with an estimated cost of £260 listed on the site R&D forms. Some radiology departments wanted to assess the cost for themselves, which led to delays of weeks. Some sites had a clear process to provide excess treatment costs, but required the trial to approach the local CCGs and help the R&D department with the application. At other sites, such as Royal Papworth Hospital, the excess treatment costs were covered directly from the radiology budget. Each site took 3 months longer than expected to open, with some sites taking in excess of 18 months.

### ***Communication between NHS trusts***

Four of the eight initial sites were PET centres, scanning patients from a number of surrounding trusts. Collection of data from trusts not directly involved in the trial proved to be extremely difficult, so the PET centres recruited patients from their own NHS trust only, which dramatically reduced their potential participant numbers.

### ***Overestimate of potential participants***

During the feasibility assessment, sites estimated the number of patients undergoing PET for nodules at their centres without paying enough attention to the inclusion/exclusion criteria.

### ***Support of chest clinicians and radiologists***

The trial was devised by radiologists, with some support from chest physicians, and each site had a radiologist as PI. However, it was vital that sites had the support of chest physicians for successful recruitment; therefore, joint PIs were appointed, with a lead radiologist and lead clinician. By engaging the support of the chest physicians, site set-up was improved, R&D approval and patient recruitment were increased.

### ***Support of multidisciplinary team co-ordinator***

At one site with poor recruitment, the SCTU met with the lung MDT co-ordinator to discuss recruitment; it was proposed that the lung MDT co-ordinator could flag patients each week by e-mail to the whole research team. This boosted recruitment and prevented loss of participants to the trial.

# Chapter 7 Economic evaluation

## Introduction

The main objectives of this chapter were to use decision-analytic modelling to assess and compare the likely costs and outcomes resulting from the use of DCE-CT, PET/CT, or a combination of DCE-CT and PET/CT as part of the management strategy for patients presenting with SPNs.

This chapter also provides a detailed description of the modelling approach, the methods and the main assumptions underpinning our chosen analysis, namely a cost-consequences analysis (CCA). The chapter continues with a section outlining the results from the main analysis and the sensitivity analyses, which identify the factors deemed to have the greatest impact on the costs and consequences of diagnostic strategies compared. The concluding section summarises the findings of the economic evaluation.

Deciding to allocate resources on the provision of diagnosing malignancy in SPNs means that these resources will be forgone and will not be available for alternative use (either to provide any other method of diagnosis of malignancy in SPNs or in the provision of different health-care interventions). The (economic) cost of this decision is the opportunity to obtain the health benefits had the resources been allocated in a different way, which, in economics, is called the opportunity cost. Economic evaluation provides decision-makers with important information about the opportunity cost of a decision. It involves the analysis of the costs and effects of one course of action, compared with alternative courses of action.<sup>52</sup>

In the current context, economic evaluation involves the assessment of costs and different types of benefits associated with alternative tests for identifying malignancy in SPNs. The form of economic evaluation adopted is a CCA. A CCA assesses alternative courses of action in terms of their impact on costs and a number of outcomes (or consequences) measured in natural or clinical units. Its aim is not to aggregate these consequences into a single measure, such as QALYs, but to highlight the trade-offs between costs and different outcomes that are implicit in a choice between the courses of action being compared. With this approach, non-health impacts can be incorporated in the analysis, something that cannot be readily achieved with QALYs and a cost-utility analysis.<sup>64</sup>

## Methods

### Model overview

As seen in the systematic review of economic evaluations (see *Chapter 5*), there are no existing economic evaluations directly comparing the use of DCE-CT and PET/CT in the diagnosis of malignancy in SPNs. Therefore, a decision-analytic model has been developed to synthesise evidence from the literature and the SPUTnik trial, and to evaluate the costs and effects of these alternative regimens. The model was designed using the software TreeAge Pro 2018, release 2 (TreeAge Software, Williamstown, MA, USA).

The model is based on a hypothetical cohort of patients aged 69 years, of which 47% are females, presenting with a SPN with a diameter of 8–30 mm. These characteristics were based on the characteristics of the sample recruited to the clinical study (see *Table 7*). Cost data were expressed in Great British pounds, reflecting prices for the year 2018. Costs were estimated from an NHS and Personal Social Services perspective. The costs and consequences of each strategy were estimated for a 2-year time horizon for the major outcomes, and projections beyond this time horizon were attempted for secondary outcomes (i.e. life expectancy and QALYs). This horizon also reflects the trial's length of follow-up.

In addition to the primary CCA, two complementary cost-effectiveness analyses were also conducted. The first of these estimated the incremental cost per malignant case treated and the second estimated the incremental cost per correctly managed case. The first analysis puts weight on correctly detecting and managing disease when it is present (i.e. it focuses on true positives). The second analysis considers that there is value in correctly managing those without a malignancy, as well as those with a malignancy (i.e. it focuses on true positives and true negatives).

The model has been built as a probabilistic model. This means that uncertainty around input parameters can be characterised by assigning probability distributions to these parameters to reflect uncertainty around the mean values. Monte Carlo simulation was used to explicitly characterise how uncertainty in the input parameters is translated into uncertainty around the results, which, in turn, were presented with their associated uncertainty (lower and upper estimates).

### Model type and rationale for model structure

The decision tree model simulates the management pathway and estimates the expected costs and consequences of a cohort of patients presenting with a SPN (of size 8–30 mm) until they are treated or until a 2-year surveillance period has been completed. The likelihood of particular courses of actions within the model is determined by a set of rules (i.e. probabilities). Different courses of action reflect alternative management pathways, which are explicitly illustrated in the decision tree. Each management pathway is assigned costs and health outcomes that are associated with events experienced by patients as they journey through the pathway. The costs, health outcomes and the likelihood of following each management pathway can be expressed as both point estimates and probability distributions.

In the primary analysis, patients entering the decision tree were assumed to be assessed for SPNs with initial diagnostic imaging using PET/CT, DCE-CT or a combination of DCE-CT and PET/CT. For the third option, individuals undergo PET/CT only if the initial DCE-CT results are indeterminate. Following the results of the imaging test, patients could be managed by immediate resection, SABR, TNB or watchful waiting (by follow-up CT). The choice of management was dependent on the outcome of the tests taken.

The structure of the decision tree was informed by models reviewed in the systematic review of economic evaluations (see *Chapter 5*) and developed in collaboration with experts in the SPutNik study. The model structure also reflected the diagnostic and management pathways outlined in the current BTS guidelines for the investigation and management of pulmonary nodules.<sup>8</sup>

The decision tree can be considered as having three main components:

1. The risk stratification component. This is the 'diagnostic component'. It stratifies patients according to the risk of having a malignant tumour, that is into groups of patients with SPNs identified as positive, negative or indeterminate for malignancy (hereafter described as 'indeterminate'). An indeterminate result could be considered a result for which clinical management decisions are taken in the face of higher uncertainty.
2. The patient management component. The second component involves patient management; it includes the procedures undertaken following an initial test result. The available management options considered after an initial test result are immediate surgical resection, SABR, confirmation by immediate TNB, watch and wait, or combinations of these options.
3. Estimating costs and outcomes. The final component of the decision tree includes the costs and outcomes for each management pathway. Short-term outcomes captured in the model include the proportion of patients managed correctly (i.e. treatment for malignant cases, and no treatment for benign cases), the proportion of malignant nodules not receiving immediate treatment (i.e. ending up in 'watch-and-wait' management) and the cost of treatment within the 2-year period. Life expectancy and quality of life are also captured to describe the long-term impact of each screening test strategy.

These three components are described in more detail in the following three subsections. They are followed by a fourth subsection, outlining the key model assumptions.

### Risk stratification component

The probability of characterising a SPN as being positive or negative for malignancy or of having an indeterminate result is conditional on the prevalence of malignancy and summary measures of the diagnostic performance of the tests estimated in the SPutNik study. Calculations for diagnostic test outcomes constitute the first part of the model, and they take the following general forms:

$$p_{\text{Positive}} = [\text{prevalence} \times \text{sensitivity}] + [(1 - \text{prevalence}) \times (1 - \text{specificity})]. \quad (1)$$

$$p_{\text{Negative}} = \text{prevalence} \times (1 - \text{sensitivity}) + (1 - \text{prevalence}) \times \text{specificity}. \quad (2)$$

While *Figure 16* shows that an imaging test result is produced in a single node with three possible outcomes (positive, negative and indeterminate), the decision model calculates this in two stages. It does this, first, by calculating the probability of a diagnostic (i.e. determinate) result, then, second, given a determinate result, by calculating the probability of a positive and negative test result, with the prevalence of malignancy updated in each step of these calculations. The estimation of the probability of a diagnostic test result uses a prevalence estimate based on a formula that is derived using the classification of the test results, as set out in *Table 17*.

The prevalence of malignancy is determined by the total number of malignant and benign cases:

$$\text{Prevalence} = \frac{\text{Malignant}}{\text{Malignant} + \text{Benign}} = \frac{g}{g + h}. \quad (3)$$

The sensitivity and specificity for a determinate test result are based on the diagnostic tests yield in patients with malignant and benign tumours:

$$\text{Sensitivity (diagnostic)} = \frac{a + c}{g}. \quad (4)$$

$$\text{Specificity (diagnostic)} = \frac{f}{h}. \quad (5)$$

The posterior prevalence of malignancy in the population with diagnostic test results is determined by applying the PLR to the pre-test prevalence, expressed as an odds ratio (OR):

$$\text{PLR} = \frac{\text{sensitivity}}{1 - \text{specificity}}. \quad (6)$$

TABLE 17 Classification of results based on test outcome and final nodule status

Test outcome	Final nodule status	
	Malignant	Benign
Cancer	<i>a</i>	<i>b</i>
Non-cancer	<i>c</i>	<i>d</i>
Indeterminate	<i>e</i>	<i>f</i>
Total cases	<i>g</i>	<i>h</i>

Worked examples of these calculations, using diagnostic outcome data from the SPutNik study, are reported in *Appendix 12*.

The probabilities of a positive or a negative test result in the population with diagnostic test results is determined in the same way, except that sensitivity and specificity are estimated using the more typical form based on the number of true-positive (TP), false-negative (FN), true-negative (TN) and false-positive (FP) test results:

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{a}{a + c} \quad (7)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} = \frac{d}{b + d} \quad (8)$$

The same procedures are applied to determine the probabilities of positive, negative or indeterminate results for each diagnostic test (PET/CT, DCE-CT and TNB).

### Patient management

Patient management options following a test result were based on the BTS guidelines for the investigation and management of pulmonary nodules.<sup>8</sup> The guidelines recommend management options including surgical excision, SABR, image-guided biopsy or CT surveillance (described as a 'watch-and-wait' strategy) depending on the risk of malignancy. These management options have been incorporated into the decision tree. The likelihood that they are offered depends on the initial result of diagnostic imaging using PET/CT or DCE-CT.

The following section provides a narrative description of the decision tree and the schematic representation in *Figure 16*. For illustrative simplicity, the tree depicted in *Figure 16* uses 'clones', which simply indicate that identical copies of parts of the management pathway are repeated in a number of places.

#### Positive for malignancy

Patients classified as positive for malignancy, by PET/CT or DCE-CT, are offered immediate surgical excision, ablative radiotherapy or a confirmatory biopsy (those cases classified as positive for malignancy at biopsy are then offered surgical excision or ablative radiotherapy). Surgical excision was also considered to have a diagnostic component based on the histopathology of excised tissue. Patients with a negative finding on histopathology were assumed to have a wedge resection without requiring a lobectomy, in comparison with those with a positive finding on histopathology, who would undergo a full lobectomy.

Patients with a negative biopsy, following a positive diagnostic imaging result, are followed-up with a watch-and-wait strategy. The watch-and-wait strategy involves repeat CT (at 3, 12 and 24 months) to assess tumour growth. Patients in whom tumour growth is indicative of malignancy are offered immediate surgical excision, ablative radiotherapy, no treatment or a confirmatory biopsy. It is assumed in the model that confirmatory biopsy during the watch-and-wait period always yields definitive diagnostic results.

Patients in whom biopsy yielded an indeterminate result, following a positive imaging test result, were offered immediate surgical excision, ablative radiotherapy, watch and wait or a second confirmatory biopsy. As above, people followed up using a watch-and-wait strategy are offered surgical excision, radiotherapy or a diagnostic confirmatory biopsy if tumour growth indicative of malignancy is observed. If a second confirmatory biopsy is offered, patients with positive and negative results are managed as outlined previously. However, if the second biopsy also yields indeterminate results, patients are offered surgical excision, radiotherapy or the watch-and-wait strategy.

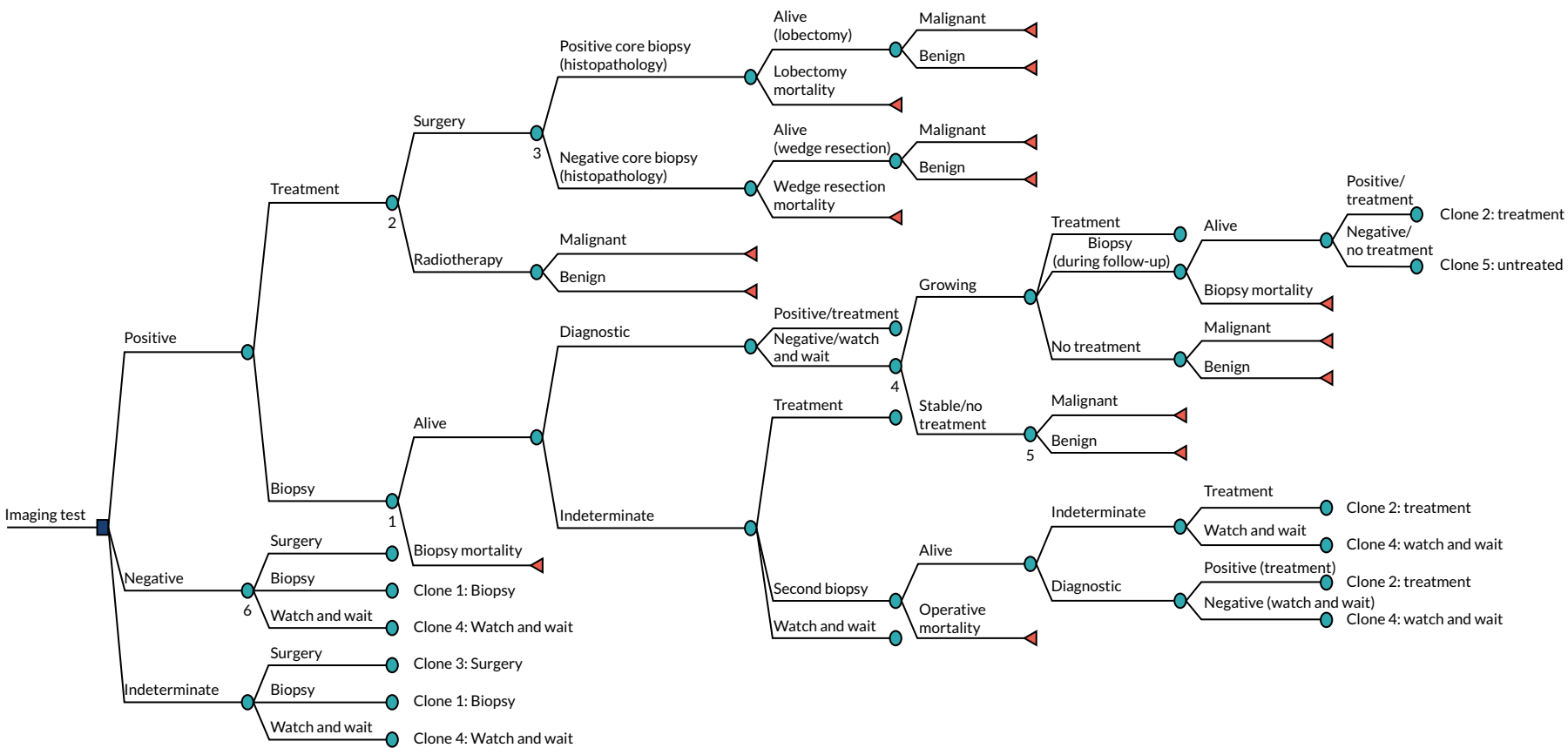


FIGURE 16 Simplified scheme of decision tree, including clones of subtrees.

**Indeterminate**

Patients whose tumour is classified as indeterminate are offered surgical excision, confirmatory biopsy or the watch-and-wait strategy. Patients offered a confirmatory biopsy or watch-and-wait strategy are offered the same options as outlined previously for those with a positive test result.

In the imaging test strategy whereby DCE-CT and PET/CT are used in combination, if DCE-CT classifies a tumour as indeterminate, PET/CT is then used as a confirmatory imaging test. In this case, the patient management decisions are based on the PET/CT results.

**Negative for malignancy**

Patients classified as negative for malignancy are likely to be predominantly entered in a watch-and-wait period. However, the decision tree has been structured to allow for patients with a negative diagnostic imaging result to be offered surgical excision or a confirmatory biopsy also. However, this is expected to occur in the minority of patients only.

**Complications and mortality associated with invasive procedures**

Mortality and non-fatal complications that may occur during surgical resection, ablative therapy or biopsy are also included in the model. To avoid excessive branching in the decision tree, non-fatal complications are implicitly included by weighting procedure-related costs and health-related utility decrements (expressed in QALYs) by the probability of non-fatal complication. Operative mortality is explicitly included in the model structure by adding a branch for procedure-related mortality after each invasive procedure.

**Model outcomes**

The current economic evaluation takes the form of a CCA, meaning that multiple outcomes are estimated in the model. In the current model, the focus is on short-term outcomes occurring during the 2-year period following the initial imaging test(s). In addition to these, secondary long-term outcomes, such as life expectancy, are also estimated and presented. A brief summary of these outcomes is presented in *Box 1*.

Model outcomes were evaluated at each decision-tree terminal node. Outcomes such as costs and utilities were summed along the path leading to each terminal node, and were dependent on the diagnostic and treatment options available in each pathway. Other outcomes, such as accuracy, correctly treated malignant tumours, and life expectancy, were calculated for each decision node and were dependent on the distribution of malignant and benign cases at these terminal nodes. The prevalence of malignancy at each terminal node following a series of diagnostic tests (PET/CT, DCE-CT, TNB) was determined by the prevalence of malignancy prior to each diagnostic test and by the sensitivity and specificity of each diagnostic test. For any diagnostic test or series of diagnostic tests, the post-test

**BOX 1** Summary of outcomes included in the decision tree

- Costs associated with the diagnostic imaging and subsequent management pathway.
- Accuracy (malignant tumours correctly treated and benign tumours left untreated).
- Malignant tumours detected and treated.
- Malignant tumours missed and left untreated.
- Malignant tumours entering the watch-and-wait strategy (receiving delayed or no treatment).
- Benign nodules inappropriately/unnecessarily treated.
- Operative deaths for patients with malignant and benign nodules.
- Life expectancy.
- QALYs.

prevalence of malignancy was estimated by applying the appropriate likelihood ratios to the pre-test prevalence, as described in *Appendix 12*. Therefore, when multiple diagnostic tests were carried out along a particular management pathway (e.g. a TNB following PET/CT), the post-test prevalence of the earlier diagnostic test (e.g. PET/CT) became the pre-test prevalence of the later diagnostic test (e.g. TNB).

### Model assumptions

To estimate the cost-consequence of the alternative imaging strategies for patients presenting with SPNs, a number of pragmatic structural assumptions had to be made:

- The process following a positive, negative or indeterminate result is identical for all screening strategies.
- Following a positive screening test result, patients are not allowed to immediately enter the watch-and-wait strategy.
- In the case of patients having two consecutive indeterminate biopsy results, a third biopsy is not an option.
- In the case of those SPNs that are under CT surveillance, biopsy always produces diagnostic results. This was carried out to confine the model's time horizon to 2 years, by preventing patients from re-entering a new CT surveillance period after an indeterminate biopsy result.
- Solitary pulmonary nodules with a negative or indeterminate imaging test result are not allowed to be treated by ablative radiotherapy without a prior confirmatory biopsy.
- Other than patients treated with ablative radiotherapy following a positive imaging test result, patients are treated with ablative radiotherapy only if they are unfit or unwilling to undergo surgery.

### Model parameters

The following sections report the parameters required to populate the model. First, clinical data are presented. The main clinical data were the prevalence of malignancy, diagnostic test accuracy and rates of procedure-related complications. Targeted searches were conducted to identify meta-analyses of diagnostic accuracy for tests included in the model that were outside the scope of SPutNIk data collection (e.g. biopsy). The systematic review of economic evaluations (see *Chapter 5*) was also used to inform clinical parameters.

In addition, unit costs and resource use associated with imaging tests and procedures captured in the model are presented. Unit costs are derived primarily from NHS reference costs<sup>65</sup> and other routine sources. Resource use assumptions for costing management strategies were based on BTS guidelines<sup>8</sup> and expert opinion from within the study team.

Finally, patient outcomes such as life expectancy and quality-of-life decrements associated with each management pathway were sourced from targeted literature searches and assumptions based on expert opinion.

### Clinical data

Patients entering the model had a starting age of 69 years. This reflected the median age of the clinical study population, as outcomes such as the prevalence of malignancy and clinical management decisions are likely to be influenced by a cohort's age. The initial prevalence of malignancy among patients presenting with a SPN with a diameter between 8 mm and 30 mm was sourced from the SPutNIk study. The prevalence estimate used in the model was 60.6%, which is similar to the prevalence reported by a previous study conducted in the UK (68.5%).<sup>22</sup> For each branch in the decision model, initial prevalence was updated while moving along that branch according to test results and diagnostic accuracy.

**Diagnostic accuracy**

*Appendix 13, Table 50*, reports the diagnostic yield, sensitivity and specificity of tests included in the model, along with the assigned distributions used in the probabilistic analysis, when applicable. The diagnostic test accuracy estimates for PET/CT and DCE-CT were derived from the trial data; for biopsy and surgery, they were derived from a review of the literature or were assumed based on expert opinion, when necessary (see *Appendix 14*). Although extrathoracic findings were also captured in the clinical study, these were not included in the analysis, as the evidence provided was not sufficient for the purpose of an economic evaluation because limited data were available on subsequent investigations and management of patients with these findings. PET/CT and DCE-CT sensitivity and specificity were calculated according to *Risk stratification component*. Because the model accounted for indeterminate tests results, the diagnostic accuracy calculations for each imaging test were based on a  $3 \times 2$  table. To account for the uncertainty of these estimates, a Dirichlet distribution was employed, with the parameters of that distribution informed by the counts of the diagnostic test results from the trial. The proportions sampled from the Dirichlet distribution were then used in the calculation of the diagnostic accuracy of each of the imaging tests.

For the diagnostic accuracy of confirmatory biopsy, the reference list of studies included in the systematic review of economic evaluations (see *Chapter 5*) was reviewed, and additional meta-analyses of biopsy diagnostic accuracy were sought. The diagnostic yield for biopsy was sourced from a recently published meta-analysis and was estimated to be 0.936.<sup>66</sup> An assumption was made that biopsy would yield the same proportion of diagnostic results for both malignant and benign nodules. A meta-analysis of the identified studies was conducted and yielded a sensitivity and specificity for lung biopsy of 0.912 (95% CI 0.872 to 0.940) and 0.955 (95% CI 0.901 to 0.980), respectively (see *Appendix 14*). The accuracy of biopsy was not assigned a probability distribution, but rather a correlation of  $-2.34$  between the sensitivity and specificity produced by the meta-analysis (see *Appendix 14*).

Surgical excision was considered to have a diagnostic component based on the histopathology of excised tissue. Histopathology was assumed to be the gold-standard method of diagnosis. This assumption has been made by previous economic evaluations,<sup>20,55</sup> and was supported by clinical experts. Patients with a negative finding on histopathology were assumed to have a wedge resection and not proceed to a full lobectomy, which was the procedure envisaged for patients with positive histopathology instead.

Growth status during watch and wait was also considered informative of the probability of malignancy, and was treated similarly to a diagnostic test. To estimate the sensitivity and specificity of the nodule's growth status, a growing nodule that was malignant was treated as a true-positive result, whereas a stable nodule that was benign was treated as a true-negative result. In addition, according to the SPUtNIk trial, 45.6% of eventually confirmed malignant tumours showed growth during the follow-up period, whereas 70.3% of eventually confirmed benign tumours remained stable. The probability of a nodule growing during the CT surveillance was estimated based on the prevalence of malignancy in patients in the watch-and-wait strategy, and the sensitivity and specificity for malignancy of nodules that showed growth or remained stable. The prevalence of malignancy at the end of the watch-and-wait period was also estimated on the basis of the sensitivity and specificity for malignancy of growing and stable nodules, by applying a likelihood ratio to the pre-surveillance prevalence of malignancy (see *Risk stratification component*).

**Procedure-related complications**

Biopsy, ablative radiotherapy and surgical resection are procedures associated with risks of complication and mortality. The identified studies from the systematic review of economic evaluations (see *Chapter 5*), as well as studies used to estimate the accuracy of diagnostic tests, were reviewed to inform the estimates for the procedure-related complications. When evidence was not available, targeted searches were conducted to identify systematic reviews and meta-analyses reporting non-fatal procedure-related complications and mortality.

Based on a targeted search for procedure-related complications (see *Appendix 15*), the results indicated that the included studies varied in the comprehensiveness and definition of reported biopsy-related complications. All included studies reported pneumothorax as an outcome, whereas more than half of the studies did not report other complications or severity of complications. As a result, the estimate of the biopsy complications rate included in the model was based on the probability of pneumothorax requiring chest tube, as this was felt to be the most clinically significant. The probability of pneumothorax requiring chest tube was reported to be 3% by a meta-analysis of 15 studies.<sup>66</sup> The mortality rate of needle biopsy (0.4%) was estimated as the weighted average mortality of patients without complications, haemorrhage and pneumothorax requiring chest tube, based on Weiner *et al.*<sup>67</sup>

Non-fatal complications of surgical procedures include atrial fibrillation, air leak, atelectasis and pneumonia (see *Appendix 15*). For the purposes of the model, we sought to identify composite estimates of the proportion of patients experiencing major non-fatal complications that would require a more intense use of resources. This was sourced from a study, included in the review of economic evaluations,<sup>55</sup> that distinguished between patients having a wedge resection and those undergoing a lobectomy (see *Appendix 13, Table 51*). It was assumed that patients undergoing surgery were at risk of wedge resection complications if the histopathology result of the surgery was negative, and at risk of lobectomy complications if the histopathology result of the surgery was positive.

According to NICE,<sup>11</sup> the most important non-fatal complication of SABR is severe cases of pneumonitis (grades 3–5). For complications of SABR, a targeted literature review was conducted because no evidence was available from the systematic review of economic evaluations (see *Chapter 5*). The search strategy for this targeted review is reported in *Appendix 15*. The non-fatal complications rate associated with ablative radiotherapy was informed by a pooled analysis of 88 studies,<sup>68</sup> for which the average reported rates of pneumonitis were 2.2% and 7.3% for grades 3–5 and grade 2 pneumonitis, respectively.

### **Clinical management following diagnostic tests**

The model includes a range of potential clinical management options following each imaging test, which include immediate intervention, biopsy confirmation or watch and wait. As clinical management decisions in the study were based entirely on PET/CT results, it was assumed that the distribution of management options following a positive, negative or indeterminate PET/CT result would apply to positive, negative and indeterminate test results from DCE-CT, or from a combination of DCE-CT and PET/CT.

Available options are partly determined by structural assumptions in the model. For example, a watch-and-wait strategy is not an option for patients with initial positive PET/CT or DCE-CT results. In the base-case analysis, the distributions across the clinical management options were based on data from the SPutNik study (see *Appendix 13, Table 52*). The likelihood of management options following indeterminate biopsy results, or an indication of a growing nodule during the watch-and-wait period, were also conditional on initial imaging test results. For instance, patients with an indeterminate imaging test result are less likely to undergo TNB than those with a positive imaging test result. To reflect uncertainty around the polychotomous transitions occurring throughout the decision tree, all clinical management parameters were assigned Dirichlet distributions (see *Appendix 13, Table 52*).

Throughout the model, it was assumed (based on trial data) that patients who were being treated with ablative radiotherapy would represent the 20% of patients who require treatment but are unfit or unwilling to undergo surgery. Based on clinical expert advice, an exception was made for the proportion of patients who were treated with SABR immediately after a positive imaging test result.

### **Resource use and costs**

#### **Costs of diagnostic investigation**

Costs for each of the two imaging tests were derived from the *NHS Reference Costs 2017/18*.<sup>65</sup> The two imaging tests included in this study are classified under the headings 'diagnostic imaging'

and 'nuclear medicine', which report costs for the named procedures categorised by other factors such as number of areas or age of subject. PET/CT was assigned a cost of £610, equivalent to the cost of PET with CT of one area. DCE-CT was assigned a cost of £133, equivalent to the cost of CT of one area with post contrast. For CT during the watch-and-wait strategy, the cost of £90 was used, equivalent to CT without contrast. This was multiplied by three, to reflect the resource use of CT during the watch-and-wait strategy.

For biopsy without non-fatal complications, the cost coded as percutaneous biopsy of lesion of lung or mediastinum was used. A weighted average of the day-case, outpatient and short-stay costs for biopsy was used according to the activity of each of the aforementioned biopsy settings (estimated as £660).

### **Cost of treatment**

The costs of surgical resection and ablative radiotherapy without complications were also sourced from *NHS Reference Costs 2017/18*.<sup>65</sup> The cost of surgery was based on the cost of complex thoracic procedures for adults without complications and was estimated to be £5778. For the cost of radiotherapy, the cost of complex conformal radiotherapy preparation was used, combined with the delivery of five fractions of adaptive radiotherapy on a megavoltage machine, summing to a cost of £2024.

### **Cost of complications**

The costs of treatment for complications associated with diagnostic or treatment procedures were also considered in the model. It was assumed that the imaging tests (PET/CT, DCE-CT, CT) were not associated with major complications that could have a significant impact on the costs and outcomes.

For surgical excision, the unit cost of complications was derived based on *NHS Reference Costs 2017/18*.<sup>65</sup> To reflect costs associated with atrial fibrillation, air leak, atelectasis and pneumonia, the average cost of complex thoracic procedures for adults without complications was subtracted from the weighted average of complex thoracic procedures for adults with moderate and major complications and comorbidities. This yielded an overall cost of £2599, which was weighted in the model with the probability of complications for wedge resection or lobectomy to calculate the expected cost of surgery-related complications.

According to a NICE guideline,<sup>69</sup> the principal non-fatal radiotherapy-related complication is moderate or severe pneumonitis. The management of pneumonitis should include the administration of prednisolone (30–40 mg) for 1–2 weeks, followed by a slow taper of the medication.<sup>70</sup> However, patients with a more severe grade of pneumonitis are expected to require prednisolone long term, and oxygen at home or hospitalisation might be required.

It was therefore assumed that patients with grade 2 pneumonitis would require a daily dose of prednisolone of 40 mg for 2 weeks, and a 40-mg dose every other day for another 6 weeks. Therefore, according to the estimated cost of 40 mg of prednisolone (£0.25),<sup>71</sup> the estimated cost of grade 2 pneumonitis was £8.87.

For patients with pneumonitis grades 3–5, it was assumed, based on Paix *et al.*,<sup>72</sup> that the administration of prednisolone and oxygen at home would be required for 3 months. The cost of oxygen at home for 3 months was estimated to be £306, based on Echevarria *et al.*<sup>73</sup> This cost included 90 days of oxygen at home, which, after being inflated to reflect current prices, was estimated to be approximately £2.50 per day. In addition, 1.5 hours of nurse time is required to set up the oxygen and provide education on its use. The hourly cost of a nurse used in the model was the average hourly cost of grades 6, 7 and 8a nurses (£45, £54, and £64, respectively).<sup>74</sup> The overall cost of grade 3 pneumonitis was estimated to be £328.

The expected costs of complications for radiotherapy were estimated by multiplying the unit costs of grade 2 and grades 3–5 pneumonitis with the associated probability of each complication presented in *Appendix 13, Table 53*.

Finally, the cost of biopsy-related complications was considered in the model. The principal complication of biopsy was pneumothorax and it was expected to result in an extended hospital admission of 1.7 days<sup>67</sup> for the vast majority of the patients experiencing this adverse event.<sup>75</sup> Therefore, the cost of an excess bed-day for pneumothorax or intrathoracic injuries with single or multiple interventions (£308) was used to estimate the unit cost for biopsy complications (£523.60).

It was assumed that patients experiencing fatal complications would require a more intense use of resources prior to death. To account for this increased resource use, and in the absence of data to inform the associated model parameters, we assumed a 20% increase in the cost of complications for patients who experienced a fatal complication.

All cost parameters were assigned probability distributions to characterise the uncertainty around their point estimates. However, for those characterised by a lesser degree of uncertainty, the mean value was used in the base-case analysis. To explore uncertainty, sensitivity analysis scenarios were conducted, whereby values for these parameters were assigned from an associated probability distribution.

### Patient outcomes

To assess whether or not substituting or combining PET/CT with DCE-CT could benefit patients if these strategies were incorporated in the management pathway, a number of patient outcomes were considered (see *Box 1*). As described in *Clinical management following diagnostic tests*, some outcomes (e.g. correctly treated cases) are evaluated at the terminal nodes of the decision tree (represented by triangles in *Figure 16*), based on the prevalence of malignancy at each of these nodes, whereas others are accumulated and calculated throughout the model. *Table 18* provides a summary of how these outcomes are calculated in the decision tree.

TABLE 18 Summary of outcome values at each terminal node

Outcome	Nodules having surgical excision or ablative therapy		Discharged without surgical excision or ablative therapy	
	Malignant	Non-malignant	Malignant	Non-malignant
Malignancy detected	1	0	0	0
Non-malignant inappropriately treated	0	1	0	0
Malignancy missed	0	0	1	0
Malignant tumours ending up in surveillance	1 (if not treated immediately)		1	
Accuracy	1	0	0	1
Operative death (benign cases)	0	1 (for fatal events)	0	0
Operative death (malignant cases)	1 (for fatal events)	0	0	0
Life expectancy	Life expectancy stage IA post resection or ablation	Normal life expectancy	Life expectancy for untreated progressive lung cancer	Normal life expectancy
Utilities	Accumulated throughout the model			
Costs (short term)	Accumulated throughout the model			

To examine how different imaging techniques could potentially affect long-term patient outcomes, we estimated the life expectancy and QALYs based on available evidence from the literature. QALYs are a combined measure of quantity and quality of life, which is expressed in terms of utility weights. *Appendix 13, Table 54*, reports the values of the life expectancy and utility parameters used in the model. Probability distributions were not assigned to these parameters, as they were treated as secondary outcomes.

### **Life expectancy**

Life expectancy for patients with different malignancy and treatment statuses was based on the exponential approximation of life expectancy method.<sup>60</sup> Under this method, it is assumed that the rate of death is constant over time and that survival is approximated by a declining exponential function:

$$S(t) = e^{-\lambda t}, \quad (9)$$

where  $S(t)$  is the probability of being alive at the period  $t$ , which measures the discrete periods passed from the commencement of the follow-up (measured in years in the current analysis), and  $\lambda$  is the hazard rate and describes the probability that patients die within the next period, given that they have survived during the current period. Based on this function, the average life expectancy can be calculated by  $1/\lambda$ , and the median life expectancy is calculated by:

$$-\ln(0.5)/\lambda. \quad (10)$$

Median life expectancies were incorporated in the model, as they are a preferred estimate of central tendency in survival analysis.<sup>76</sup>

Life expectancy for people with benign nodules not experiencing fatal complications was based on UK life tables.<sup>77</sup> The survival rate for a population aged 69 years was used and the median life expectancy was estimated to be 17.41 years.

Life expectancy for patients with malignant tumours treated with ablative radiotherapy or surgical excision was taken from a UK cohort study of patients with presumed stage I cancer undergoing treatment with curative intent.<sup>78</sup> The median life expectancy in this cohort was estimated to be 8.83 and 3.99 years for patients treated with surgical excision (i.e. lobectomy) and ablative radiotherapy, respectively. It was assumed that patients with benign nodules undergoing surgical excision (i.e. wedge resection) or ablative radiotherapy would have a life expectancy of 17.41 years if they had not experienced a fatal complication. Patients with malignant nodules who were left untreated were assumed to have a life expectancy of 1.38 years. This estimate was based on a cohort of 1693 untreated operable stage I to stage III non-small-cell lung cancer patients.<sup>79</sup>

### **Utilities**

A targeted search of evidence was conducted to identify utility values to inform the model parameters. Utility values reflecting the quality of life of patients with SPNs and the utility decrements associated with procedure-related complications were sparse. Only one study included in the review of economic evaluations had reported utility data.<sup>20</sup> However, a more recently published study was identified,<sup>72</sup> which formed the basis of informing the utility parameters in the model. The utility weights used in the model are reported in *Appendix 13, Table 54*.

The utility of people with benign nodules was assumed to be equivalent to that of the general population. This was estimated based on Kind *et al.*,<sup>80</sup> and was 0.78 for people aged 65–74 years and 0.73 for those aged  $\geq 75$  years. The utility value for patients with malignant cancer who were treated with surgical excision or ablative radiotherapy was based on the utility of patients with advanced metastatic non-small-cell lung cancer,<sup>81</sup> and was estimated to be 0.712. Finally, for the utility of patients with untreated malignant nodules, an assumption (based on Paix *et al.*)<sup>72</sup> was made that this utility would be equivalent to that of treated patients experiencing symptoms of dyspnoea and pain. This was estimated to be 0.421.

Patients undergoing procedures were assumed to experience complications, which would have an impact on their quality of life. Consequently, a utility decrement was applied to patients experiencing non-fatal complications. The utility decrement parameters were sourced from Paix *et al.*<sup>72</sup> The utility decrement for patients experiencing a biopsy-related complication was assumed to be 0.011 per month. It was assumed that this decrement would be applied for only 1 month after the biopsy complication. For patients experiencing radiotherapy-related complications, a utility decrement of 0.03 was assumed for the first 6 months after treatment, and 0.0775 for the following 6-month period, resulting in a total composite utility decrement of 0.108. Finally, for patients with surgical complications, a total utility decrement of 0.042 over the first 6 months was considered. It was assumed that the utility decrement from wedge resection-related and lobectomy-related non-fatal complications was identical.

## Analysis

### Base-case analysis

Outcomes of the base-case analysis were presented in the form of a CCA. Model outcomes associated with each imaging strategy were tabulated and presented separately as summary estimates.

In addition, two joint estimates of the included outcomes were presented, as they were considered to provide meaningful information. These were (1) the incremental cost per correctly treated malignancy and (2) the incremental cost per correctly managed case. These additional outcomes provide information on what additional resources are required by either <sup>18</sup>F-FDG-PET/CT alone or <sup>18</sup>F-FDG-PET/CT combined with DCE-CT, compared with DCE-CT alone, to correctly identify and manage a malignant case or to correctly manage either a malignant or a benign case. The joint estimates of costs and effects were estimated in an incremental analysis and were presented in the form of ICERs for each comparator. In the absence of a decision rule [i.e. a willingness-to-pay (WTP) threshold], a cost-effectiveness acceptability curve was produced to reflect how uncertainty in the input parameters influences the result of the economic evaluation. This is achieved by presenting the probability of a strategy being cost-effective for a range of WTP thresholds.

### Sensitivity analyses

To capture how uncertainty in the input parameters is translated into uncertainty in the model outputs, both deterministic and probabilistic sensitivity analyses were used. The majority of parameters included in the model were assigned probability distributions to describe the degree of uncertainty around the parameter mean value. The choice of the appropriate distribution for each parameter was based on the properties of each parameter. For instance, cost parameters were assigned a gamma distribution to reflect that costs must have a positive value. Different types of distributions were characterised on the basis of the different parameters' mean value and standard deviation and the shape of the statistical distribution. Monte Carlo simulation was then used to assign sampled values from the probability distributions to the input parameters. This assignment process was repeated 1000 times to form a distribution of values for each model outcome. To ensure replicability of the values sampled from the probability distributions, a pseudorandom number generator was used for the sampling of values from distributions assigned to each model parameter. The results of Monte Carlo simulations are generated and reviewed in the form of a cost-effectiveness scatterplot.

For parameters that had not been assigned a probability distribution, deterministic sensitivity analyses were performed to explore the impact of each parameter or group of parameters on the model outputs. A description of scenarios explored as part of the sensitivity analyses is provided in this section.

### Exploratory univariate sensitivity analyses

To identify the parameters with the highest impact in relevant model outcomes, multiple univariate sensitivity analyses were conducted. This involved varying the range of all parameter values by 50%, when possible. The impact of the model parameters on specific model outputs was depicted using tornado diagrams. Based on the results of this initial sensitivity analysis, the 15 parameters with the

highest impact on the incremental costs and correctly managed cases were presented. Results and conclusions from other outcomes were narratively described.

### ***Scenario in which indeterminate results are not allowed in the model***

To reflect how decisions on patient management are made based on imaging test results and clinical interpretation, a result of an imaging test in the base-case analysis could be indeterminate. To test the impact of indeterminacy, a sensitivity analysis was conducted to explore a scenario in which the tumours of patients in the model could not be classified as indeterminate for malignancy, and such patients were instead treated in the same way as patients with a positive imaging test result. Consequently, in the absence of indeterminate results, the imaging strategy of DCE-CT followed by PET/CT (when DCE-CT results are indeterminate) becomes equivalent to using DCE-CT alone. Hence, in this scenario analysis, DCE-CT was compared with PET/CT only.

To estimate the accuracy of the imaging tests for this analysis, two different approaches were taken. In the first approach, the diagnostic accuracy of the tests was derived from the radiologist classification of SPNs from the trial data. Based on these, patients with indeterminate test results were assumed to be treated as positive cases. In the second alternative approach [(alt)], the diagnostic accuracy of each test was based on pre-defined thresholds set in the trial, described in *Chapter 6, Diagnostic accuracy of positron emission tomography-computerised tomography and dynamic contrast-enhanced computerised tomography*. Summary estimates of imaging test accuracy used in this analysis can be found in *Appendix 16, Table 62*.

### ***Model validation***

The developed decision-analytical model was subjected to a number of validity checks. First, the structural assumptions were presented to clinical experts and other stakeholders with expertise in the area and good knowledge of the decision problem. The model went through an iterative process of refinements until it was agreed that all the principal aspects of the decision problem were captured.

Throughout the model development, different tests were conducted to ensure that the model was producing plausible results and behaved as expected to a different number of changes in the inputs. This was carried out by testing whether or not costs and outcomes were accumulating as expected throughout each pathway, and whether or not the proportion of the cohort that was reaching each terminal node was as anticipated, given the findings of past studies and expert clinical opinion.

Finally, a between-model consistency test was undertaken. The model was compared with another model that was independently developed by two other members of the team to answer the same decision question. Throughout the development of the model, continuous collaboration with these members was established, with the aim of ensuring a common understanding of the decision problem and consistency in the input parameters. Overall agreement was reached across both models, with minimal discrepancies.

## **Results**

### ***Base case***

*Table 19* presents the results of the CCA. On average, DCE-CT alone was the least costly strategy, and the difference in costs between PET/CT alone and a combination of DCE-CT and PET/CT was small, on average, and may lack economic significance. However, a combination of DCE-CT and PET/CT was the best alternative strategy when comparing patient outcomes. This strategy correctly identified the highest proportion of patients with malignant disease (46.7%), had the lowest proportion of malignant cases left without treatment (13.71%) and achieved the lowest proportion of inappropriate treatment in patients with a benign nodule (9.0%). These results led to the appropriate management of 84.4% of patients, according to their malignancy status.

TABLE 19 Costs and consequences results for base-case analysis

Single outcomes	PET/CT (SD)	DCE-CT (SD)	DCE-CT and PET/CT (SD)
Cost (£)	4013 (206)	3305 (199)	4058 (210)
Accurately managed cases (%)	82 (1.6)	77.8 (2)	84.4 (1.4)
Malignancies treated (%)	44.2 (2.5)	40.1 (2.5)	46.7 (2.4)
QALYs	7.64 (0.25)	7.43 (0.26)	7.76 (0.24)
Life expectancy (years)	10.5 (0.32)	10.22 (0.34)	10.65 (0.31)
Delayed or no treatment (%)	20.31 (1.84)	25.63 (2.15)	17.18 (1.59)
Malignancies missed (%)	16.2 (1.68)	20.49 (2.04)	13.71 (1.43)
Benign cases treated (%)	9.91 (1.25)	9.8 (1.25)	9 (1.23)
Operative deaths (%)	1 (0.05)	0.92 (0.05)	1.05 (0.05)
Operative deaths for benign cases (%)	0.17 (0.02)	0.16 (0.02)	0.15 (0.02)
Operative deaths for malignant cases (%)	0.96 (0.05)	0.87 (0.05)	1.01 (0.05)

SD, standard deviation.

In addition, PET/CT, on average, performed better than DCE-CT in reducing the malignant cases receiving late or no treatment. However, when the two tests were used in combination, the proportion of malignant cases having delayed or no treatment was reduced by a further 3.13% (20.31% vs. 17.18%). DCE-CT was associated with the lowest life expectancy, on average (10.22 years), and fewest QALYs (7.43), of all the strategies, as a result of the larger number of patients with malignant nodules left untreated, which reduces a patient's life expectancy and quality of life. A combination of DCE-CT and PET/CT was associated with the highest life expectancy (10.65) and QALYs gained (7.76) per patient.

Finally, the combination of DCE-CT and PET/CT was associated with the lowest proportion of operative deaths for patients with benign nodules (0.15%), because of the higher overall test specificity. However, the same strategy was associated with an increased proportion of operative deaths in all patients with SPNs, compared with DCE-CT (1.05% vs. 0.92%, respectively), because of the larger number of patients who are offered procedures with associated fatal complications.

Results for the incremental cost per malignant case treated and incremental cost per correctly managed case are presented in *Table 20*. PET/CT is excluded from *Table 20*, as it was an extendedly dominated strategy (i.e. a combination of DCE-CT and DCE-CT plus PET/CT strategies would be a more efficient strategy than PET/CT alone). The cost-effectiveness acceptability curves for the incremental cost per malignant case treated and incremental cost per correctly managed case are presented in *Figures 17* and *18*, respectively.

*Figure 17* shows the probability of each strategy being cost-effective for a wide range of maximum WTP-per-malignant-case-treated thresholds. When the WTP ceiling ratio per correctly treated malignancy was < £9000, DCE-CT was always the most preferable strategy. However, when the WTP ceiling ratio for one more correctly treated malignancy increased to > £15,500, the strategy that combines DCE-CT and PET/CT became the most cost-effective, with a probability of 1.

TABLE 20 Cost-effectiveness results for base case analysis

Strategy	Cost (£)	Incremental cost (£)	Effectiveness (%)	Incremental effectiveness (%)	ICER
<i>Per malignant case treated</i>					
DCE-CT	3305		40.1		
DCE-CT and PET/CT	4058	753	46.7	6.61	11,395
<i>Per correctly managed case</i>					
DCE-CT	3305		77.8		
DCE-CT and PET/CT	4058	753	84.4	6.65	11,323

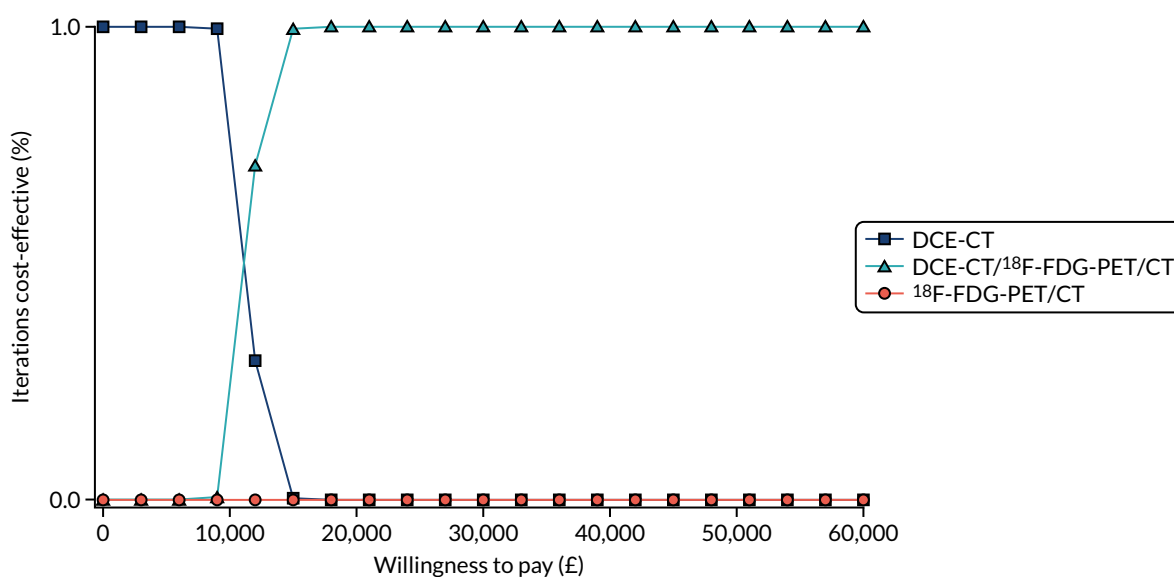


FIGURE 17 The cost-effectiveness acceptability curve for cost per correctly treated malignancy.

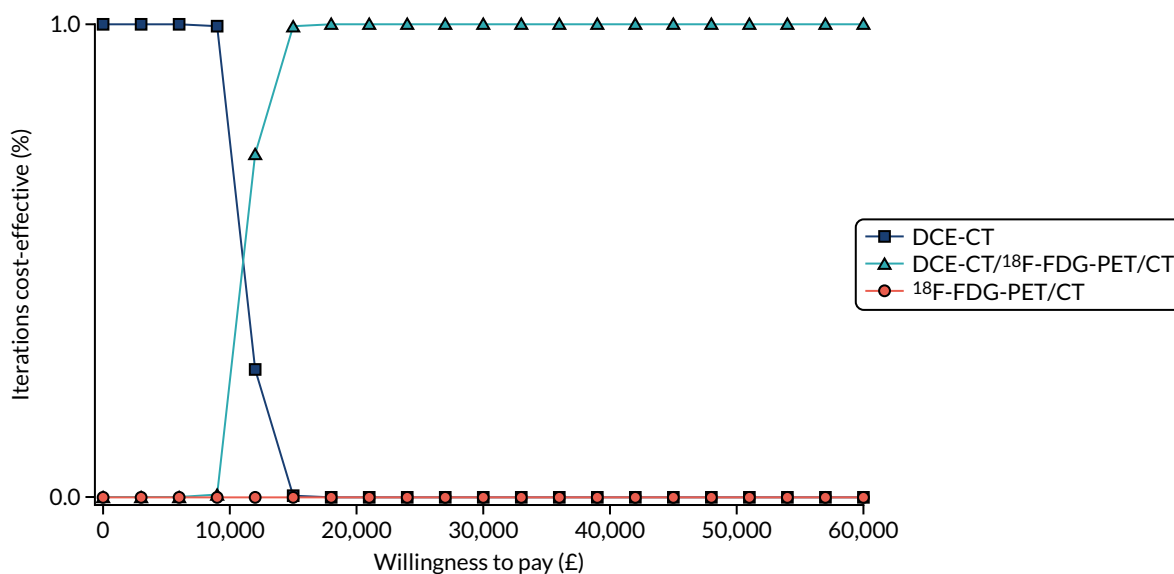


FIGURE 18 The cost-effectiveness acceptability curve for cost per correctly managed case.

Another analysis (see *Figure 18*) on the incremental cost per correctly managed case took into account also those benign cases who were correctly provided no treatment. In this analysis, DCE-CT was more likely to be the preferred strategy when the WTP ceiling ratio per correctly managed case was < £11,395. However, when the WTP ceiling ratio per additional correctly managed case was increased beyond £16,000, DCE-CT combined with PET/CT became the strategy most likely to be considered cost-effective.

## Sensitivity analyses

### Exploratory univariate sensitivity analyses

A univariate deterministic sensitivity analysis was conducted to identify the parameters with the highest impact on costs and on the proportion of correctly managed cases. Since the base-case results showed that PET/CT was an extendedly dominated strategy, the impact of the parameters on the incremental costs and incremental correctly managed cases were presented for DCE-CT and DCE-CT combined with PET/CT.

The results from the univariate sensitivity analysis showed that the main drivers for the cost difference between the DCE-CT and DCE-CT plus PET/CT strategies are the initial prevalence of malignancy, the accuracy of PET/CT and the costs associated with PET/CT and surgical excision (*Figure 19*). The results also showed that, when the sensitivity of biopsy drops (depicted with a blue bar), the cost difference between the two strategies decrease, substantially. The cost of procedures, such as biopsy, radiotherapy and CT, had a moderate impact on the cost difference between the DCE-CT and DCE-CT plus PET/CT strategies. The cost of complications had no impact on the results. Overall, varying the model parameters over a range of 50% did not make DCE-CT a more costly strategy than DCE-CT plus PET/CT.

The impact of changing the model parameter values by 50% was also estimated on the effectiveness of the diagnostic strategies when defined by the proportion of correctly managed patients (*Figure 20*). The univariate sensitivity analysis showed that the accuracy of PET/CT and biopsy, the diagnostic yield of DCE-CT in benign cases and the initial prevalence of malignancy had the greatest impact on the incremental effectiveness between DCE-CT and DCE-CT plus PET/CT. Other parameters, such as the proportion of patients with a growing nodule being treated, the proportion of malignant nodules showing growth over the surveillance period and the likelihood of undergoing surgery immediately after the imaging test results, all had a moderate impact on effectiveness. Overall, this sensitivity analysis showed that the sensitivity of PET/CT was the only parameter that, if reduced by 50%, would make DCE-CT the more effective strategy.

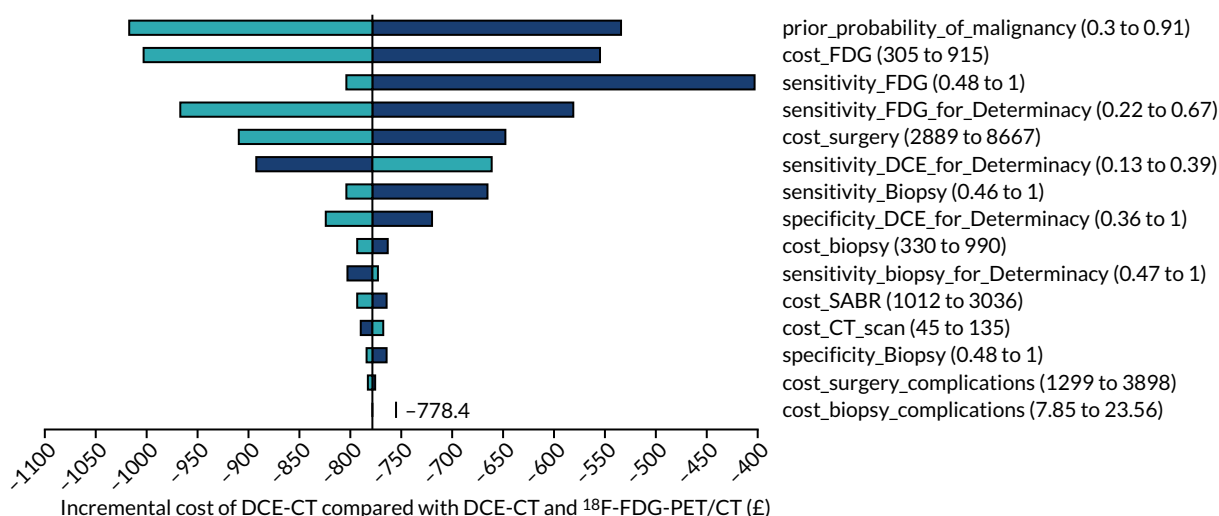


FIGURE 19 Tornado diagram showing impact on incremental costs of changes in key parameters.

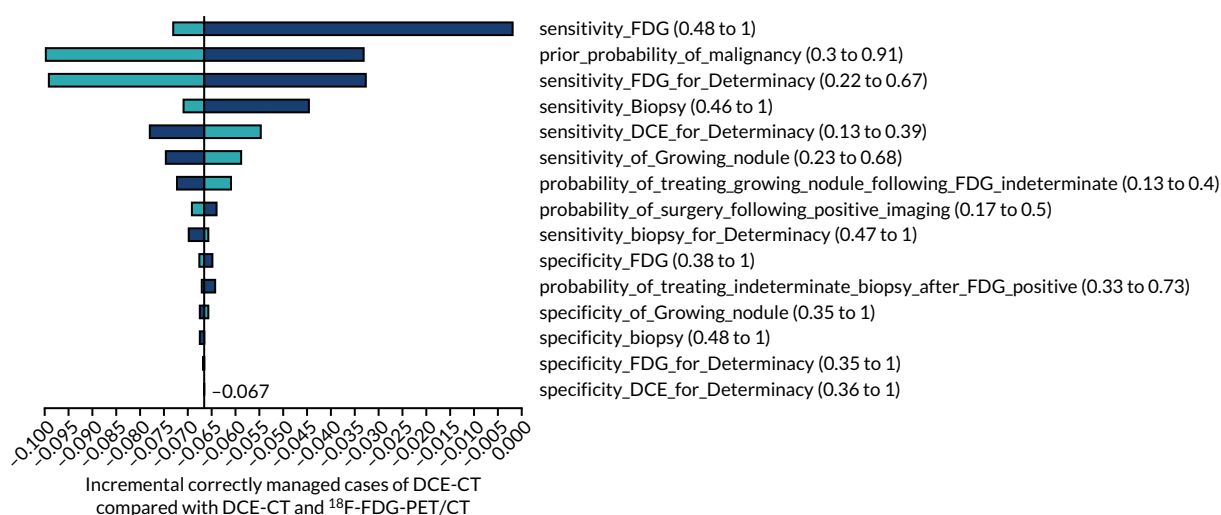


FIGURE 20 Tornado diagram showing the impact on the incremental cost per correctly managed case.

In total, the diagnostic accuracy of PET/CT, the initial prevalence of malignancy, and the diagnostic yield of DCE-CT in benign cases were the main drivers of the results when comparing DCE-CT with DCE-CT plus PET/CT. This was also the case when comparing the effectiveness of these strategies in terms of correctly treated malignancies and QALYs. In addition, when the two strategies were compared in terms of total QALYs, the results were also sensitive to the life expectancy following a lobectomy and the quality of life of benign and treated malignant cases.

**Scenario in which indeterminate results are not allowed in the model**

For this analysis, costs and consequences of four imaging test alternatives, which are assumed to always yield diagnostic test results, were estimated. In this scenario, PET/CT and DCE-CT are based on the diagnostic accuracy of those tests where indeterminate cases are treated as positive. PET/CT (alt) and DCE-CT (alt) are based on the diagnostic accuracy as derived from a pre-determined set of thresholds.

As Table 21 shows, PET/CT (alt) was, on average, the least costly strategy (£3957), whereas PET/CT was the most costly strategy (£4945). In contrast, PET/CT (alt) was the imaging strategy associated with the least favourable outcomes when considering the number of correctly managed patients, patients with malignant

TABLE 21 Costs and consequences results of scenario excluding indeterminate results

Single outcomes	PET/CT (SD)	DCE-CT (SD)	PET/CT (alt) (SD)	DCE-CT (alt) (SD)
Cost (£)	4945 (287)	4469 (287)	3957 (275)	4358 (279)
Accurately managed case (%)	92.8 (0.8)	92.8 (0.8)	85 (1.9)	92 (0.9)
Malignancies treated (%)	55.8 (2.7)	55.8 (2.7)	47.4 (3.1)	54.9 (2.7)
QALYs	8.26 (0.22)	8.26 (0.22)	7.75 (0.26)	8.2 (0.23)
Life expectancy	11.3 (0.29)	11.3 (0.29)	10.63 (0.33)	11.22 (0.3)
Delayed or no treatment (%)	4.65 (0.74)	4.65 (0.72)	15.98 (2.16)	5.95 (0.94)
Malignancies missed (%)	3.95 (0.69)	3.95 (0.68)	13.35 (2)	5 (0.89)
Benign cases inappropriately treated (%)	15.06 (2.16)	15.08 (2.16)	6.36 (1.67)	14.08 (1.97)
Operative deaths (%)	1.29 (0.06)	1.29 (0.06)	1.05 (0.07)	1.26 (0.06)
Operative deaths, benign cases (%)	0.25 (0.03)	0.25 (0.03)	0.11 (0.02)	0.23 (0.02)
Operative deaths, malignant cases (%)	1.2 (0.06)	1.2 (0.06)	1.02 (0.07)	1.18 (0.06)

SD, standard deviation.

tumours receiving delayed or no treatment, and missed malignancies. When accounting for operative death for benign cases and the proportion of benign cases inappropriately treated, PET/CT (alt) was associated with the most favourable outcomes, because of the high specificity compared with the other diagnostic test strategies.

Overall, PET/CT and DCE-CT yield more favourable results with regards to patient outcomes than the other two strategies. PET/CT and DCE-CT achieve almost identical patient outcomes, with DCE-CT costing, on average, £476 less than PET/CT.

## Discussion

The results of the CCA suggest that using DCE-CT alone is likely to reduce the overall costs related to the imaging strategy and management pathway of people presenting with a SPN (of 8–30 mm in diameter). However, current practice (i.e. PET/CT) was associated with better outcomes, including, but not limited to, the accuracy of treatment according to malignancy status, as well as the life expectancy and quality of life of patients with SPNs. Furthermore, the analysis suggests that the combined use of the two imaging techniques is likely to further improve outcomes for a minimal additional cost, compared with current practice. Multiple univariate one-way sensitivity analyses showed that the prevalence of malignancy and the sensitivity of PET/CT are the model parameters with the highest impact on incremental costs and incremental correctly managed cases.

The results of this economic evaluation should be interpreted with caution, as assumptions had to be made about the management of patients. In particular, owing to the limited evidence concerning the management of patients following a DCE-CT or DCE-CT plus PET/CT test result, it was assumed that these patients were managed identically to patients with the equivalent PET/CT test result. That is, the model assumes that a positive test result provides the same information to guide patient management regardless of how it was made and that patients who receive a true-positive result do not differ regardless of how a diagnosis is made. In addition, extrathoracic findings, which could potentially increase the clinical value (as well as the costs) associated with <sup>18</sup>F-FDG-PET/CT, and the costs associated with the treatment of these findings, were not included in the analysis.

The choice of model structure and the limited evidence on utility weights following a patient's diagnosis and management mean that the estimated QALYs associated with each imaging test strategy should not be considered robust. It is partly for this reason that an incremental cost per QALY has not been estimated (a further reason is that we did not estimate costs beyond 2 years). Hence, further research is required to obtain more evidence for model inputs, particularly for those related to the management following DCE-CT or DCE-CT plus PET/CT, and long-term outcomes such as life expectancy and quality of life.

The structure of the decision-analytic model is a simplified representation of the management of patients with SPNs following the results of an imaging test. The model was used to synthesise evidence from multiple sources of evidence, with the SPUTNIK trial being the primary source of evidence. There is uncertainty surrounding both the specific model inputs and the underlying structure of the model. Hence, important costs and benefits may not be accurately estimated. Nevertheless, considerable efforts were made to ensure that the best available evidence was used and that the model reflects real-life decisions (i.e. inclusion of indeterminate imaging test results) and the management pathway to the greatest extent.



# Chapter 8 The IPCARD-SPN questionnaire substudy: the identification of patient-elicited symptoms that predict malignant pulmonary nodules – an exploratory analysis

## Background

The SPUtNik study participants were invited to complete a self-administered symptoms questionnaire [Identifying symptoms that Predict Chest And Respiratory Disease (IPCARD-SPN)]<sup>82</sup> at the point of recruitment to the SPUtNik study (see the SPUtNik study protocol<sup>39</sup>). The questionnaire (see the SPUtNik study protocol<sup>39</sup>) was designed to record a broad range of chest, respiratory and systemic symptoms that might predict lung cancer, and it had demonstrated good content validity, test-retest reliability and completion rates in validation studies.<sup>82</sup> The IPCARD-SPN questionnaire was developed in a population with early-stage lung cancer and includes items designed to elicit symptoms often normalised by lung cancer patients prior to diagnosis.<sup>83</sup>

## Aims

- To identify symptoms that predict malignant SPNs in a population with indeterminate SPNs.
- To determine whether or not the inclusion of symptoms found to distinguish between malignant and non-malignant nodules increases the diagnostic value of DCE-CT and PET/CT.

## Methods

### *Recruitment and questionnaire completion*

Potential SPUtNik participants were invited to complete the IPCARD-SPN questionnaire prior to attending the imaging appointment and to bring it along to the appointment. At the appointment, the research nurse or radiographer took consent for the SPUtNik study, which included an optional part for the IPCARD-SPN substudy. If a patient consented to the IPCARD-SPN substudy, completed questionnaires were left with the research nurse, with an option of returning incomplete questionnaires by post within 2 days. A second copy of the IPCARD-SPN questionnaire was posted to patients who had not received a lung cancer diagnosis after 12 months (range 12–18 months) of follow-up.

### *Data analysis*

The predictive value of symptoms for lung cancer diagnosis was identified at 12 and 24 months following recruitment. Participants who did not receive a lung cancer diagnosis in year 1 and participants who were < 18 months post recruitment at the time when the study amendment was introduced were invited to complete a second IPCARD-SPN questionnaire. Owing to low eligibility and recruitment rates for the second questionnaire, completion rates were too low to enable analysis of the second questionnaire. Missing IPCARD-SPN questionnaire data for participants included in the SPUtNik main analyses also meant that it was not possible to ascertain whether or not the inclusion of symptoms found to distinguish between malignant and non-malignant nodules increases the diagnostic value of DCE-CT and PET/CT.

## Recoding

Items indicating the presence or absence of generic symptoms (discomfort in chest, upper body or shoulders; a cough for > 3 weeks; breathing changes; unexpected tiredness; coughing up blood; sweats) were recoded to the binary categories 'experienced currently or within the previous 3 months' or 'experienced > 3 months ago or never', except haemoptysis, which was recoded to ever/never. Symptom descriptor item categories indicating frequency (never, once, occasionally, most of the time) were recoded to a binary variable ever/never. Items indicating when a symptom started were recoded to generate two response categories: 'within the previous 3 months' or '> 3 months ago'. Items measuring how much discomfort or distress a symptom caused on a 10-point scale not found to have any discriminant value in previous validation studies were excluded from the analyses.

## Model development

Logistic regression was used to explore associations between each of the IPCARD-SPN questionnaire symptom variables and lung cancer, at 1 year and 2 years following recruitment, adjusted for age and sex. Symptoms associated with lung cancer at  $p \leq 0.05$  were selected for model entry.

The multivariable model was developed in a sample of the data set that had complete data for the entry variables, to avoid missing data determining variable selection.

Forward stepwise multiple logistic regression analysis was used to enter symptom variables into the analysis that were associated with lung cancer in the univariate analysis ( $p \leq 0.05$ ). The criterion for entry into the model was  $p \leq 0.05$  and the criterion for removal was  $p$ -value of  $\geq 0.1$ . To reassess variables discarded through the stepwise process, each was added in turn to the final model to identify any improvement in model fit [assessed using the Akaike information criterion (AIC), with a lower score indicating a better fit]. The adjusted ORs for symptoms were reported with  $p$ -values. The variables included in the final model were entered into a multivariable regression model in a subsample of the data set with complete data for those variables. As some variables had been dropped during the forward stepwise process, this sample was larger than the sample in which the model had been developed. Adjusted ORs in this larger sample were presented for comparison.

All multivariable models were adjusted for age, sex and smoking history: variables that might be expected to account for differences in symptoms between those with and those without lung cancer. Smoking history data from the IPCARD-SPN questionnaire were more complete and had a stronger association with lung cancer than comparable data extracted from participants' medical records for the SPUtNIk study, and were used in model development. Common comorbidities that share symptoms with lung cancer (chronic respiratory disease and cardiovascular disease), extracted from participants' medical histories in secondary care notes, were investigated for associations with lung cancer diagnosis. Comorbidities that differed between lung cancer and non-lung cancer were included as covariates in the multivariable models. Chronic obstructive pulmonary disease (COPD) precedes lung cancer diagnosis in 40–90% of the cases, and differences in the prevalence of COPD between those with and those without lung cancer might explain differences in symptoms between these groups.<sup>84</sup> However, missing data for COPD status (60%) meant that it was not possible to adjust for this covariate in analyses.

Previous research has found that interactions between lung cancer symptoms and risk factors have higher PPVs than symptoms alone.<sup>85</sup> Established epidemiological risk factors for lung cancer (e.g. asbestos exposure and previous malignancy<sup>86</sup>) and the clinical risk factor 'nodule size at baseline scan' were extracted from participants' medical notes. Asbestos exposure had a high number of missing data (82.6%) and was not included in the analyses. Interaction terms were added to final models to test for effect modification with age, sex, years smoked, previous malignancy and nodule size at baseline.

## Model accuracy

The area under the curve with 95% CI was calculated for the models predicting lung cancer diagnosis at year 1 and year 2. To identify the threshold (cut-off point) that best distinguished between those

with and those without lung cancer, sensitivity, specificity and Youden's index were calculated with 95% CIs for predicted probability values of 0.1–0.9.

Predicted probabilities for a male participant and female participant of average age and average number of smoking years were calculated for symptoms independently associated with lung cancer at year 1, keeping all of the other symptoms absent.

### Checking assumptions

Linearity assumptions between the logit and the continuous variables were checked in the final models by treating continuous variables as categorical variables and checking for any substantial change in coefficients and standard errors of the other variables in the model. Standardised residuals were plotted against fitted values and residuals for age and smoking plotted against fitted values to check for outliers. There was no evidence of extreme outliers. The variance inflation factor was used to test for multicollinearity in the final models.

## Results

### Return and data completion rates

The first IPCARD-SPN questionnaire was returned with a valid participant ID for 287 participants. Thirty-nine of these participants were not included in the SPUtNIK study main analyses. The second questionnaire was returned for 66 participants, 56 of whom had also returned the first questionnaire. A total of 281 participants had IPCARD-SPN (questionnaire 1) and lung cancer outcomes data and were included in the analyses exploring associations between symptoms and lung cancer diagnosis (see *Appendix 17, Table 63*, for participant characteristics). A total of 56.5% (156/281) and 58.7% (165/281) of participants received a lung cancer diagnosis during the first 12 and the first 24 months of surveillance, respectively.

Missing data for symptom variables ranged from 2% to 22%. Missing data for covariates included in model development ranged from 3% (age and sex) to 13% (years smoked).

### Covariates and interactions

The continuous variable 'years smoked' was associated with lung cancer in univariate analyses (1.03,  $p < 0.0001$  at year 1; and 1.024,  $p < 0.0001$  at year 2). Although age and sex did not have statistically significant univariate associations with lung cancer (OR 1.02,  $p = 0.28$ , and OR 0.70,  $p = 0.14$ , for age and sex, respectively, at year 1; and OR 1.02,  $p = 0.13$ , and OR 0.68,  $p = 0.12$  for age and sex, respectively, at year 2), they have been associated with lung cancer and symptoms in previous studies<sup>82,83</sup> and were included in all multivariable models.

The binary clinical covariates 'any cardiovascular disease' and 'any respiratory disease', used in the main SPUtNIK analyses, were not strongly associated with lung cancer at year 1 (OR 1.30,  $p = 0.39$ , and OR 1.41,  $p = 0.17$ , respectively) or at year 2 (OR 1.61,  $p = 0.12$ , and OR 1.25,  $p = 0.37$ , respectively), when adjusted for age and sex, and were not entered into the multivariable models. For COPD, 60% of the data were missing, and it had a non-significant relationship with lung cancer (OR 1.69;  $p = 0.23$ ).

The epidemiological risk factor 'previous malignancy' and the clinical risk factor 'nodule size at baseline scan' were associated with lung cancer at year 1 (OR 2.56,  $p = 0.017$ , and OR 1.08,  $p = 0.003$ , respectively) and at year 2 (OR 3.40,  $p = 0.003$ , and OR 1.08,  $p = 0.002$ , respectively), when adjusted for age and sex, and were entered into the multivariate models.

## Symptoms associated with lung cancer at year 1

### Single-symptom analyses

The OR and  $p$ -values for the symptom variables associated with lung cancer at  $p < 0.05$ , adjusted for age and sex, are reported in Table 22 (lung cancer diagnoses within 1 year). The symptom descriptor 'A niggle, ache or pain that feels like wind or indigestion but not associated with eating' and the symptom variables 'Unexpected tiredness first experienced within the previous 3 months' and 'More colds or flu in the previous 12 months' were positively associated with lung cancer at  $p < 0.05$  at year 1; 'sweats not caused by the menopause within the previous 3 months' was negatively associated with lung cancer. The four variables associated with lung cancer in the single-symptom analyses were entered into the multivariable model.

### Multivariable analyses

The ORs of variables independently associated with lung cancer at year 1 ( $p \leq 0.05$ ) and the AIC of the model, once all excluded variables had been checked against the final model, are reported in Table 23.

TABLE 22 Odds ratios and  $p$ -values for symptoms associated with lung cancer at year 1 ( $p \leq 0.05$ ), adjusted for age and sex

Symptom	OR (95% CI)	$p$ -value
A niggle, ache or pain that feels like wind or indigestion but not associated with eating (reference category: symptom absent)	1.85 (1.02 to 3.34)	0.041
Unexpected tiredness first experienced in the previous 3 months (reference category: first experienced > 3 months ago or never)	2.64 (1.06 to 6.53)	0.036
More colds or flu in the previous 12 months (reference category: no more colds or flu in the previous 12 months)	2.097 (1.04 to 4.21)	0.037
Sweats not caused by the menopause in the previous 3 months (reference category: no sweats)	0.292 (0.085 to 1.002)	0.05
Sweats not caused by the menopause > 3 months ago (reference category: no sweats)	0.934 (0.504 to 1.733)	0.831

TABLE 23 Multivariable models for lung cancer diagnosis at year 1 (model 1: entry of variables significant at 0.05)

Model for lung cancer diagnosis at year 1	Model 1			
	Model development sample with complete data for all entry variables ( $n = 181$ )		Sample with complete data for variables that remained in the final model ( $n = 227$ )	
	OR (95% CI)	$p$ -value	OR (95% CI)	$p$ -value
Intercept	0.53	0.644	0.424	0.477
Sex (reference category: female)	0.53 (0.27 to 1.04)	0.065	0.544 (0.30 to 0.97)	0.042
Age	1.003 (0.96 to 1.04)	0.861	1.009 (0.97 to 1.04)	0.615
Years smoked	1.035 (1.01 to 1.05)	0.00006	1.029 (1.01 to 1.04)	0.00009
Sweats not caused by the menopause in the previous 3 months (reference category: no sweats)	0.06 (0.005 to 0.602)	0.017	0.171 (0.34 to 1.40)	0.018
Sweats not caused by the menopause > 3 months ago (reference category: no sweats)	0.70 (0.32 to 1.50)	0.360	0.699 (0.34 to 1.40)	0.316
Unexpected tiredness first experienced in the previous 3 months (reference category: first experienced > 3 months ago or never)	5.09 (1.27 to 20.22)	0.021	3.44 (1.20 to 9.91)	0.022
AIC	231.4		295.8	

Two of the symptom variables associated with lung cancer in the univariate analyses ( $p \leq 0.05$ ), 'a niggles, ache or pain that feels like wind or indigestion but not associated with eating' and 'more colds or flu in the previous 12 months', and the risk factors 'nodule size at baseline' and 'previous malignancy' did not remain in the multivariable model. The variables included in the final model were also entered into a multivariable regression model in a subsample of the data set with complete data ( $n = 227$ ) for those variables (see *Table 23*). All symptom variables remained independently associated with lung cancer ( $p \leq 0.05$ ) and the ORs changed little. Interactions were investigated between symptoms and other model covariates (nodule size, previous malignancy, age, sex, years smoked), but did not improve model fit.

### Model accuracy

The area under the curve for model 1 (*Figure 21*) was 0.708 (95% CI 0.641 to 0.775). The optimum threshold for distinguishing between those with and those without lung cancer for model 1 that provided the best balance between sensitivity and specificity was  $c = 0.5$  [with a specificity of 0.61 (95% CI 0.51 to 0.71), a sensitivity of 0.70 (95% CI 0.61 to 0.78) and a Youden's index of 0.31 (95% CI 0.12 to 0.49)]. Not restricting the sample to observations for which there were complete data during model development led to models with greater areas under the curve. However, variable selection might then be determined by missing data, rather than strength of association with lung cancer. See *Appendix 17, Table 64*, for the predicted probabilities for the diagnosis of lung cancer within 12 months of surveillance (model 1).

The symptoms and risk factor variables discarded from model 1 (previous malignancy, nodule size at baseline scan, 'a niggles, ache or pain that feels like wind or indigestion but not associated with eating' and 'more colds or flu in the previous 12 months') were added, in turn, to model 1 in the larger sample with complete data for the model 1 variables ( $n$  ranged from 204 to 224). Independent associations with lung cancer approached, but did not reach, statistical significance ( $p \leq 0.05$ ) for three of these four variables (*Table 24*).

### Symptoms associated with lung cancer at year 2

#### Single-symptom analyses

The ORs and  $p$ -values for all symptom variables associated with lung cancer at  $p < 0.05$ , adjusted for age and sex, are reported in *Table 25* (lung cancer diagnoses within 2 years). The symptoms 'feeling out of breath' and 'taste changes in the previous 2 years' were positively associated with lung cancer at year 2 ( $p \leq 0.05$ ).

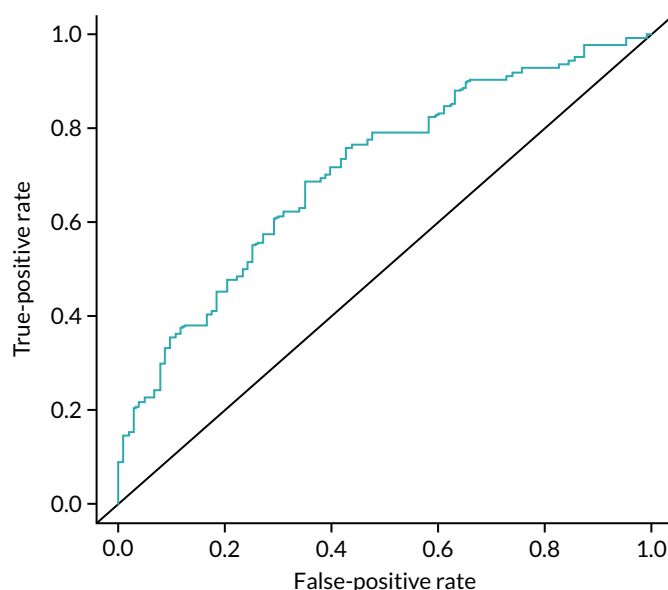


FIGURE 21 The ROC curve for model 1.

TABLE 24 Independent associations with lung cancer for variables re-entered into model 1<sup>a</sup>

Risk factor/symptom	n	OR (95% CI)	p-value
Previous malignancy	224	2.7 (1.13 to 6.45)	0.025
Nodule size at baseline scan	222	1.06 (0.99 to 1.12)	0.055
More colds and flu in the previous 12 months	210	2.25 (0.93 to 5.43)	0.071
A niggle, ache or pain that feels like wind or indigestion but not associated with eating	204	2.08 (0.97 to 4.47)	0.06

a Variables were re-entered using a sample with complete data for the final model 1 variables.

TABLE 25 The ORs and p-values for symptoms associated with lung cancer at year 2 ( $p \leq 0.05$ ), adjusted for age and sex

Symptom	OR (95% CI)	p-value
Feeling out of breath	1.95 (1.12 to 3.38)	0.017
Experienced taste changes in the previous 2 years	2.57 (1.26 to 5.22)	0.009

When forward stepwise regression was used to select symptom variables and risk factors associated with lung cancer at year 2 for inclusion in a multivariable model, the symptom variables did not improve model fit and did not remain in the model (Table 26). On removal of the risk factor variables 'previous malignancy' and 'nodule size at baseline scan' from the model, the symptom variables were still not independently associated with lung cancer and did not improve model fit.

### Model accuracy

The area under the curve for the model predicting lung cancer at year 2 (Figure 22) was 0.72 (95% CI 0.653 to 0.782). The optimum threshold for distinguishing between those with and those without lung cancer for model 2 was  $c = 0.5$ , with a sensitivity of 0.79 (95% CI 0.71 to 0.85), a specificity of 0.51 (95% CI 0.40 to 0.61) and a Youden's index of 0.29 (95% CI 0.11 to 0.46).

TABLE 26 Multivariable model for lung cancer diagnosis at year 2

Model for lung cancer diagnosis at year 2	Model 2			
	Sample with complete data for entry variables (n = 212)		Sample with complete data for variables remaining in the model (n = 235)	
	OR (95% CI)	p-value	OR (95% CI)	p-value
Intercept	0.076	0.050	0.081	0.045
Sex	0.633 (0.338 to 1.18)	0.153	0.653 (0.364 to 1.17)	0.152
Age	1.019 (0.98 to 1.05)	0.298	1.014 (0.98 to 1.04)	0.428
Years smoked	1.031 (1.01 to 1.04)	0.00008	1.029 (1.01 to 1.04)	0.0001
Previous malignancy	3.611 (1.34 to 9.67)	0.011	3.985 (1.58 to 10.004)	0.003
Nodule size at baseline scan	1.057 (0.99 to 1.11)	0.054	1.075 (1.01 to 1.13)	0.011
AIC	267.02		295.55	

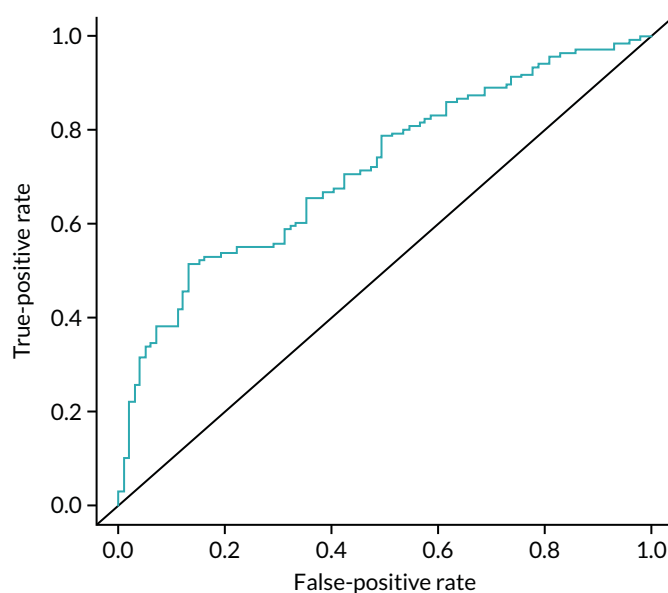


FIGURE 22 The ROC curve for model 2.

## Discussion

The completion of the IPCARD-SPN questionnaire by SPuTNIK study participants provided the opportunity to identify symptoms that might predict early lung cancer in a population under surveillance for SPNs; as far as we are aware, the potential for symptomatic diagnosis has not yet been investigated in this population. The IPCARD-SPN questionnaire was designed to include a broad range of symptoms and lay symptom descriptors that might distinguish between lung cancer and non-malignant diagnoses in populations with differing spectrums of disease and comorbidities.<sup>82</sup> As symptoms that predict lung cancer might be expected to differ between populations with different spectrums of disease,<sup>87</sup> the IPCARD-SPN questionnaire does not provide a score, but elicits a range of potential predictors that might then be used to generate population-specific risk scores.

Exploratory analyses identified four symptoms associated with lung cancer diagnosis within the first year of surveillance. When forward stepwise regression analyses were used to select symptoms for a multivariable model, two of these four symptoms, 'sweats in the previous 3 months, not caused by the menopause' and 'unexpected tiredness first experienced in the previous 3 months', remained in the model [adjusted for age, sex and years smoked (model 1)]. The risk factors 'nodule size at baseline scan' and 'previous malignancy', although associated with lung cancer in univariate analyses, did not improve model fit, and therefore exited the model. Symptoms associated with lung cancer diagnosis within 2 years of surveillance in single-symptom analyses (adjusted for age and sex) were not independently associated with lung cancer and did not improve model fit once 'years smoked' was added to the model. The addition of 'previous malignancy' and 'nodule size at baseline scan' improved the performance of the model predicting lung cancer diagnosis at year 2 (model 2).

### *Clinical plausibility and confounding*

'Sweats in the previous 3 months, not caused by the menopause' had a negative association with lung cancer at year 1. Sweats are considered to be a symptom of late-stage lung cancer and would not be expected to have a positive association with malignant pulmonary nodules. Sweats and tiredness might be associated with chronic chest and respiratory disease. As those with lung cancer had smoked for longer than those without lung cancer, and smoking is a risk factor for other chest and respiratory diseases that share common symptoms with lung cancer, it is possible that symptoms that predict lung cancer in our model are caused by smoking or smoking-related chronic diseases that are more prevalent in those with malignant pulmonary nodules. Missing data for quantity of tobacco use and

clinical covariates meant that it was not possible to adjust for pack-years smoked or COPD. Imputation of missing smoking history data in future models, and prospective and systematic identification of COPD status, might better identify symptom predictors that are independent of smoking history and common smoking-related disease. However, if the occurrence of new symptoms that predict lung cancer in patients under surveillance of indeterminate SPNs was to be fully explained by COPD or other smoking-related disease in future research, this would be important information to impart to patients at the start of a surveillance protocol.

### Limitations

Prospective data collection was uncertain when, owing to missing data, it was not possible to ascertain that the questionnaire had either been completed at the site prior to imaging, or returned by the participant within 2 days of recruitment; these questionnaires might have been completed following results that indicated an increase or decrease in the likelihood of lung cancer. However, the single-symptom analyses ( $p \leq 0.05$ ) identified a small number of symptom predictors.<sup>8</sup> If symptoms that differed between the lung cancer and non-lung cancer groups had been due to non-prospective data collection (that is, participants who knew that they had or were more likely to have lung cancer were more likely to interpret bodily sensations as symptoms and complete the questionnaire accordingly), then differential reporting might be expected to apply to multiple potential lung cancer symptoms that are prevalent in a SPN surveillance population. The large number of symptom variables included in the IPCARD-SPN questionnaire also increases the likelihood of type 1 error during the model development phase. In univariate analyses, 4 of 74 symptom variables were associated with lung cancer at year 1 and 2 of 74 symptom variables were associated with lung cancer at year 2 ( $p = 0.05$ ). Given the number of statistical tests, it is possible that the univariate associations were explained by type 1 error. Symptoms identified as predictors of lung cancer diagnosis in these models are presented as candidate symptoms for malignant pulmonary nodules only, and further research will be required to validate the model.

A priori sample size calculations suggested that 257 cases of lung cancer and 118 participants without lung cancer would provide 80% power to detect a difference of 16% in a common symptom (from 34% in non-lung cancer to 50% in lung cancer). As the analyses were underpowered, it is possible that predictors with low and moderate effect sizes were missed. The four variables that exited model 1, 'a niggles, ache or pain that feels like wind or indigestion but not associated with eating', 'more colds or flu in the previous 12 months', nodule size at baseline scan and previous malignancy, were added, in turn, to model 1 in a sample with complete data for the model 1 variables (a larger sample than the sample used for model development:  $n$  ranged from 204 to 224).  $p$ -values ranged from 0.025 for previous malignancy ( $n = 224$ ) to 0.071 for 'more colds or flu in the previous 12 months' ( $n = 210$ ). It is possible that these four variables would be independently associated with lung cancer in a larger sample, and might be considered as candidate predictors in future research.

### Conclusion

The development cohort provided by the SPUTNIK study has enabled the identification of candidate symptom predictors for the diagnosis of lung cancer within the first 12 months of the surveillance of indeterminate pulmonary nodules; these symptom predictors can now be further investigated in fully powered studies that systematically record respiratory and cardiovascular comorbidities.

# Chapter 9 Discussion

## Main findings

### Study conduct

This pragmatic multicentre prospective study was conducted in 16 UK hospital sites and recruited patients with indeterminate SPNs. The median age was 69 years, and 75% were current or ex-smokers, with 21% having had a previous exposure to asbestos, silica or coal. A prevalence of malignancy of 68.5% was used in the original sample size calculation and our overall rate of 61% was in keeping with other diagnostic accuracy studies of PET/CT or DCE-CT in SPNs.<sup>22</sup> There was some variation in cancer prevalence across the recruiting sites, with the majority lying between 50 and 80%, although there was also a large variation in the number of participants that the sites recruited. This probably reflects variation in the use of PET/CT in the diagnostic pathway between secondary and tertiary care centres, the latter often having an 'enriched for malignancy' referral population. It is standard practice to follow up lesions that are thought to be benign with regular CT and not to subject patients to unnecessary biopsy; this pathway was followed for 77% of the benign lesions.

Both diagnostic tests were rigorously monitored by central QA teams and repeatability of interpretation was within acceptable limits, but reflected the breadth of performance across NHS sites. For PET/CT, when a SUVmax of > 2.5 was used as the threshold for malignancy, there was excellent agreement between the site read and core laboratory read, but, in visual grading, the sites were more likely to rate a nodule as having greater than mediastinal signal than the central read. The central read was not as good at CT morphology, as only the low-dose CT scan was available; this does not have as clear features as normal routine CT scans, and had high interobserver variability.

For DCE-CT, using the maximum enhancement of  $\geq 20$  HU as malignant, there was good agreement between the site read and core laboratory read. There were, however, wide LOAs between the two measures, suggesting substantial variability in the technique. Further work is required to determine if this is due to differences in analysis software or scanners, or due to variations in the size and precise location of each of the ROIs at each of the time points. Most sites were unfamiliar with this technique and were not using it in routine practice. At some sites, the DCE-CT scan was read by the same readers as the PET/CT scan, rather than by dedicated chest radiologists. A more uniform approach may emerge if this technique is clinically adopted, especially if dedicated software incorporating the appropriate analysis workflow is used and if the examination is read by thoracic-trained radiologists.

### Study results

As expected, the largest subtype of lung cancer was non-small-cell carcinoma (81%), with 74% of those being adenocarcinoma. The remainder were as follows: carcinoid, 7%; small cell, 4%; and other, 8%. This is consistent with other studies of SPNs and is reflective of the types of cases being investigated in the UK.

The imaging technique with the highest sensitivity is DCE-CT (see *Figure 18*), which achieves a value of 96.8% when using an enhancement cut-off point of  $\geq 15$  HU and 95.2% when using an enhancement cut-off point of  $\geq 20$  HU, whereas PET/CT achieved a sensitivity of only 72.5%. However, this increased sensitivity comes at the penalty of low specificity: DCE-CT achieves specificities of only 20.3% and 29.3% at enhancement cut-off points of  $\geq 15$  HU and  $\geq 20$  HU, respectively, compared with specificity of 80.5% for PET/CT.

Among the 312 patients with matched examinations, the sensitivity, specificity and diagnostic accuracy were 72.5%, 80.5% and 75.6%, respectively, for PET/CT and 95.2%, 29.3% and 69.2%, respectively, for DCE-CT when using the PET and CT grading and the maximum enhancement of  $\geq 20$  HU. In this study,

the PET/CT sensitivity was lower and the specificity higher than in many other studies. It is well documented that adenocarcinoma can have a low FDG uptake, compared with other cancers, and the high prevalence of this histological subtype in this study may account for the lower PET/CT sensitivity. In a 2018 meta-analysis of 20 studies of SPNs,<sup>12</sup> the pooled sensitivity was 89% (95% CI 87% to 91%) and the pooled specificity was 70% (95% CI 66% to 73%). When comparisons were made according to nodule size (8–15 mm, just over 15 mm to 20 mm, and just over 20 mm to 30 mm), there was increasing sensitivity of PET/CT with increasing nodule size, a corresponding drop in specificity and an increase in diagnostic accuracy from 74.9% to 81.5% at a pre-defined SUV cut-off point of 2.5. There was an optimal threshold of SUV of > 1.9 for smaller nodules (8–15 mm in size), with specificity greater than sensitivity. The meta-analysis did not examine the effects of SPN size on sensitivity and specificity, and our study may have had relatively smaller SPNs than other published results, which would account for these results. The changes seen with nodule size on PET/CT is due to partial volume effects as a result of the lower resolution of the PET examination; however, resolution is improving with the next generation of scanners.

Tumour type can affect the sensitivity of PET/CT. Adenocarcinoma can have a low uptake of FDG, resulting in a negative test. Tumour grade and FDG SUV have been well correlated and the SUV can give independent prognostic information in stage 1 lung adenocarcinoma.<sup>88</sup> Analysis by type and grade of tumour is planned in future work, as it may be that DCE-CT is more suited to the detection of adenocarcinoma.

The meta-analysis of DCE-CT included 23 studies with 2397 participants and demonstrated a pooled sensitivity and specificity of 95% and 76%, respectively. Although our cohort study matches the high sensitivity of 95% using an enhancement threshold of 20 HU, we could not match the published literature with our specificity, which was much lower at 29%. The source of the poor specificity in the current study is not clear. The closest prior study to our own is that by Swensen *et al.*,<sup>15</sup> who examined the diagnostic accuracy of DCE-CT in 356 patients across seven sites, finding a sensitivity of 98% and specificity of 58% using a threshold of 15 HU. That both our own prospective multicentre study and that of Swensen *et al.*<sup>15</sup> find a much lower specificity than that suggested by the meta-analysis speaks to the significant impact of the inherent bias of retrospective studies performed in single academic centres. Even this prior multicentre study was highly selective in its selection of suitable nodules, excluding all those with 'signs of necrosis, cavitation, calcification, or a low signal-to-noise ratio'.<sup>15</sup> When compared with the largest study,<sup>89</sup> which had a less stringent selection criteria and included 486 patients with solid or part-solid nodules, the sensitivity and specificity were 98% and 46%, respectively (with malignancy in 114 of 249 nodules), again coming far closer to our own experience. Thus, it may be that, in academic centres, using stringent selection criteria, higher specificity can be achieved. However, the purpose of the current study was to examine how DCE-CT would perform in as close to a real-world environment as possible. Further sources of the poor specificity that must be considered are wide variation across sites in interpretation, as mentioned previously, or differences in software. Reporting at sites allowed a three-point scale of positive, indeterminate and negative, and this was used in the health economic analysis. This uncertainty is reflected in the poor specificity found. We further analysed the results by SPN size and found there was relatively little change in the sensitivity or specificity with increasing nodule size, although there was a slight improvement in diagnostic accuracy.

Exploratory modelling was undertaken, looking at various parameters at different thresholds. This demonstrated that the SUVmax was the most diagnostically accurate measurement, with an AUROC curve of 0.86; this increased if it was combined with DCE-CT peak enhancement.

### Health economics

Despite the poorer performance of DCE-CT, compared with PET/CT, the markedly lower cost of the examination results in a potentially more cost-effective strategy for the investigation of a SPN. To model cost-effectiveness of using DCE-CT first and, if positive, then using PET/CT, pre-defined thresholds were created. This showed sensitivity, specificity and diagnostic accuracy of 69.8%,

82.1% and 74.7%, respectively. When society's WTP per correctly treated malignancy was < £9000, DCE-CT was always the preferable strategy. However, when society's WTP to gain one more correctly treated malignancy increased to £16,000, the strategy that combines DCE-CT with PET/CT becomes the strategy most likely to be considered cost-effective, with a probability equal to one. The availability of CT across the NHS is much greater than that of PET/CT. PET/CT is commissioned by NHS England via a national contract that covers the wide geographical area, but is delivered in certain centres only. The reliability of FDG radiotracer supply can be problematic, resulting in postponed examinations, which adversely affects patients. PET/CT cannot normally be performed on the same day as the diagnostic CT because of the logistics of tracer and scanner availability. However, this has to be balanced by the huge demands on CT machines: often, there is a reporting backlog and reports can take several weeks. Finally, the health economic assessment is heavily dependent on the costs of the diagnostic tests. In the UK, the costs of PET/CT dropped in 2018, which would influence the model and, other things being equal, would improve the cost-effectiveness of using PET/CT, compared with other strategies which do not involve PET/CT.

## Strengths and limitations

To our knowledge, this is the largest multicentre prospective trial to date comparing DCE-CT with PET/CT in the diagnosis of indeterminate SPNs in the same patients. We have shown that DCE-CT has higher sensitivity and that <sup>18</sup>F-FDG-PET/CT has higher specificity and better accuracy. Sensitivity for <sup>18</sup>F-FDG-PET/CT is lower for nodules of < 20 mm, whereas DCE-CT is unaffected by nodule size. A major strength of this study is that it was pragmatic, reflecting real-life practice, and the comparison was conducted in the same cohort of patients. Patients were recruited from accredited centres from NHS hospitals across the UK in a prospective manner before diagnosis in a standardised setting, and are representative of the type of patients with an undiagnosed SPN. The average age, smoking status, presence of COPD and historical exposure to asbestos, coal or silica were similar to those seen in patients referred for investigation through outpatient clinics and lung cancer MDT meetings.

From a technical perspective, a broad range of acquisition parameters for the imaging tests were accepted, reflecting current NHS practice, and were similar to those that would be used if this technique was widely adopted. Our study used nodule diameter as the entry criteria, but clinical practice has moved towards volumetric measurement to determine which diagnostic pathway to follow. However, the size criteria we used are still relevant. In this study the radiation dose for DCE-CT was higher, at 30 mSv, than a standard PET/CT dose of 14.1 mSv. The DCE-CT dose was high to reduce noise and improve the sensitivity of the technique. However, the high radiation dose of DCE-CT would be a concern and would need to be reduced before adoption in routine clinical practice.

From the health economics perspective, a major strength is that the modelling was based on an updated systematic review of the current evidence linked to standard NHS costs and applied to results from this study. The exploratory analysis benefits from a large well-curated data set with robust outcome data, from which it has been possible to derive new meaningful cut-off values for different sizes of nodules. However, it is recognised that these findings need to be validated in a separate unrelated data set.

One limitation of the study was that the specificity of 29% obtained by DCE-CT was significantly lower than the 75% that the meta-analysis of the literature would suggest. The frequently poor/indeterminate quality of the included studies in the meta-analysis raises the possibility that significant biases were present, inflating the actual accuracy of the technique. The other possibility is that the large number of centres in the current study included many with relatively limited experience of the technique, with reporting often being undertaken by non-thoracic radiologists, which could, in turn, lead to a more conservative assessment, with overcall, rather than undercall, of the likelihood of malignancy. This is only the second multicentre analysis of DCE-CT, to our knowledge, with the other analysis being performed in centres with almost a decade of experience developing and subsequently reporting the technique.<sup>15,90</sup>

Future exploration of techniques to reduce this in-built reader variability and impact of experience are warranted. Standardisation of the technique with automated/semiautomated contouring of the nodules or volumetric analysis of the nodules to account for the effects of breathing changing the relative position of the ROI are both avenues worthy of further research. Another shortcoming was the low recruitment rates and the length of time it took to recruit patients, mainly as a result of governance and regulatory changes in the R&D departments delaying set-up of the additional sites. However, the study remains relevant and the examinations use current techniques. Finally, the reported 61% malignancy rate of nodules in the current study is relatively high. This is representative of solitary nodules of this size detected in clinical practice, and is in keeping with previous meta-analyses of MRI and PET in SPNs;<sup>12,50</sup> it is substantially higher than the malignancy rate of screening-detected SPNs, such as in the National Lung Screening Trial (15.0% malignancy in nodules of 10–30 mm)<sup>91</sup> and the Dutch–Belgian Lung Cancer Screening (NELSON) trial (15.2% malignancy in nodules of > 10 mm).<sup>92</sup> Previous work has shown the sensitivity of a technique to be relatively robust to disease prevalence and for the specificity to increase with falling prevalence.<sup>93</sup> It can be postulated that the diagnostic accuracy of DCE-CT would be similar, or even further improved, for SPNs found in a screening population. However, this requires further evaluation.

Regardless of these limitations, the study remains relevant and the examinations use current techniques that can be easily applied in everyday clinical practice.

The clinical predictors of unexpected tiredness first experienced in the previous 3 months, and more colds or flu in the previous 12 months were positively associated with lung cancer, which is highly novel and shows some interesting results, but it is emphasised that these are exploratory findings and of little value until demonstrated in a prospective study on a different data set.

## Lessons for the future

It is very resource intensive to undertake follow-up over a 2-year period following recruitment. This is particularly hard for the research nurses who are often charged with this task. With the advent of NHSX, all digital NHS data will be available from one legal entity. As information on outcomes and resource use is now collected from electronic patient records, it should be possible to create an electronic CRF that can pull data into the study database. Obviously, this can be carried out for routinely collected data only, but this could go a long way towards reducing trial costs, and the challenges of identifying personnel at each site to extract data from patient records and enter it into another database. Attempting to pilot this approach in this ethics-approved study, in which patients had consented to have their records examined, was not possible because of governance concerns raised from several quarters. Streamlined flow of patient data into clinical trial databases from electronic patient records and central NHS databases would greatly facilitate many clinical trials.

## Future research

In recent years, several major lung cancer screening studies have shown that screening high-risk populations with low-dose CT confers an overall mortality benefit.<sup>94,95</sup> The randomised US National Lung Screening Trial of three annual low-dose CT examinations in high-risk smokers showed a 20% reduction in mortality, compared with chest radiography,<sup>94</sup> which has been sustained on extended follow-up.<sup>96</sup> The NELSON<sup>97</sup> population trial of regular CT screening in people aged 50–74 years who are at high risk of cancer from smoking showed a reduction in mortality over a 10-year period of 26% in men and 39% women, compared with a control arm. In this trial, there was a 0.9% cancer detection rate, with 50% found at stage I.<sup>97</sup> As a result, CT screening programmes are being developed in a number of countries. In the UK, following a successful pilot study in Manchester that specifically targeted screening at current smokers and those of low socioeconomic status, NHS England has recently set up targeted lung health checks at 14 pilot sites in 10 schemes across England to evaluate

CT screening in high-risk populations.<sup>5,98</sup> It is hoped that these pilots will, in due course, lead to the roll-out of a national screening programme in England.

With the advent of screening, the number of patients with a SPN requiring further investigation will increase substantially. A previous HTA review<sup>6</sup> noted that CT screening is associated with a relatively high false-positive rate and subsequent investigations constitute a significant cost. Furthermore, SPNs are a common finding in whole-body screening CT examinations offered to asymptomatic individuals by independent-sector providers. Typically, the costs of follow-up investigations from these examinations are incurred in the public sector (UK National Screening Committee, appendix 17).

The findings from the SPUtNik study raise the issue as to how DCE-CT could be used in the context of a screening programme. Given that the prevalence of lung cancer in a high-risk screened population (0.9–3%) is much lower than in the SPUtNik study, it is recommended that PPV and NPV calculations for both DCE-CT and PET/CT are undertaken with this lower prevalence, especially with smaller lesions. As DCE-CT costs less, it could be tested as part of a two-step screening procedure whereby a SPN is found and confirmed by an algorithm and a DCE-CT examination is undertaken at the same visit as the screening CT. Validation of the new SUV and enhancement thresholds for SPNs according to size would be worthwhile, and combining this with radiomic information from the CT and PET examinations might improve accuracy. Further work is required to reduce the radiation dose of DCE-CT without compromising the sensitivity.

Further work on the model is warranted. The internal validity of the final model was assessed by the bootstrap resampling technique to adjust for overoptimism in the estimation of model performance due to validation in the same data set that was used to develop the model itself. It is important to test the validity of the model in an external, unseen data set of SPNs.

Exploration of an integrated examination with DCE-CT performed with the PET examination would be a very efficient way to combine information from both the techniques at a reduced examination cost. This could be piloted to estimate clinical effectiveness and cost-effectiveness.

In summary, the research recommendations are as follows:

- explore the integration of the DCE-CT component into the PET/CT examination for the characterisation of SPNs
- explore the feasibility of two-stage CT lung screening with DCE-CT at the same visit if a suspicious nodule is found
- undertake analysis of PET/CT and DCE-CT by tumour type, grade and size using different SUVs and enhancement thresholds to improve accuracy
- develop a new protocol for DCE-CT with a lower radiation dose suitable for the newer CT machines.



## Chapter 10 Conclusion

In conclusion, this study has shown that, although PET/CT is more accurate, DCE-CT is likely to reduce the overall costs. The case for adopting the less costly DCE-CT would be strengthened if it could be incorporated efficiently into CT lung screening workflows in the future, but only if the radiation dose is reduced. New optimal cut-off values for PET SUVs according to the size of the nodule could improve the performance of PET/CT. The combination of SUVmax and peak enhancement had the greatest AUROC curve for nodule discrimination. A combined PET/CT plus DCE-CT approach is likely to further improve accuracy for a minimal additional cost, compared with current practice.



# Acknowledgements

The trial was funded by the NIHR HTA programme. It was co-ordinated by SCTU, which is directed by Professor Gareth Griffiths and is part-funded by Cancer Research UK.

This work was undertaken at University College London Hospital/University College London, which received a proportion of funding from the Department of Health and Social Care's NIHR Biomedical Research Centres funding scheme.

## Ethics committee

The SPUtNik trial was approved by the South West Research Ethics Committee. It received ethics approval on 30 July 2012.

## Sponsor

University Hospital Southampton NHS Foundation Trust sponsored this trial and approved the original protocol and subsequent amendments to the study.

## Trial Steering Committee members

- Dr Toby Maher (Consultant Respiratory Physician, Chairperson).
- Dr Simon Padley (Consultant Radiologist).
- Professor Jannet Dunn (Statistician).
- Mrs Lisa Lamond (PPI member).
- Professor Stephen Duffy (Statistician).

## Data Monitoring Committee members

- Mr Seth Seegobin (Chairperson).
- Professor Willie Hamilton.
- Professor Vicky Goh.

## Collaborators

We would like to thank Jeremy Jones for his contribution to designing the study, developing the health economic model and writing report chapters.

### *SPUtNik investigators*

Lesley Gomersall, Jonathan Bennett, David Baldwin, Kristopher Skwarski, John O'Brien, Steve O'Hickey, Nick Adams, Andrew Scarsbrook, Richard Riley, John Buscombe, Neal Navani, Sarah Doffman, Kenneth Jacob, Joris van der Horst, Joseph Sarvesvaran, Ravi Sharma and Rajiv Srivastava.

### *SPUtNik research nurses*

Theresa Green, Amanda Stone, Kathleen Collie, William Hickes, Sarah Goodwin, Patricia Clark, Louise Nelson, Kathryn Moore, Amy Gladwell, Beena Poulouse, Alison Porges, Robert Anderson, Victoria Ashford-Turner, Maria Machado, Dawn Thornton, Harvey Dymond, Jayne Tyler, Raquel Gomez,

Susan Mbale, Gail Pottinger Andrea Lodge, Robert Shortman Sue King, Elaine Smith, Sandra Beech, Barbara McLaren, Jane Lyttle, Hugh Lloyd-Jones, Anne Joy and Tania Pettit.

### *SPUtNik radiographers*

Elizabeth Robertson, Claire Napier, Diane Lowe, Jan Bush, Georgina Haywood, James Hunter, Alison Fletcher, Nick Weir, Clare McKeown, Mary Dempsey, Joanne Wormleighton, Garry McDermott, Elizabeth Crawford, Julie Turkas, Kerry Edwards, Paul Holland, Gabrielle Azzopardi, Paul Murphy, Richard Smith, Leigh Clements, Julie Butler, Rebecca Dillon, Elizabeth Llewellyn and Juttalie Cole.

## Contributions of authors

**Fiona J Gilbert** (<https://orcid.org/0000-0002-0124-9962>) (Professor, Honorary Consultant Radiologist and Co-Chief Investigator) was involved in the design of the study, delivery of the study, interpretation of the results and writing of the report.

**Scott Harris** (<https://orcid.org/0000-0001-5774-1537>) (Associate Professor of Medical Statistics) was involved in the design of the study, delivery of the study, statistical analysis, interpretation of the results and writing of the report.

**Kenneth A Miles** (<https://orcid.org/0000-0002-5920-381X>) (Honorary Professor, Radiology and Nuclear Medicine) was involved in the design of the study, delivery of the study and interpretation of the results.

**Jonathan R Weir-McCall** (<https://orcid.org/0000-0001-5842-842X>) (University Lecturer, Radiology) was involved in the delivery of the study, interpretation of the results and writing of the report.

**Nagmi R Qureshi** (<https://orcid.org/0000-0003-2674-449X>) (Consultant Cardiothoracic Radiologist) was involved in the design of the study, delivery of the study and interpretation of the results.

**Robert C Rintoul** (<https://orcid.org/0000-0003-3875-3780>) (Reader in Thoracic Oncology) was involved in the design of the study, recruitment of participants, delivery of the study and interpretation of the results.

**Sabina Dizdarevic** (<https://orcid.org/0000-0002-6715-3744>) (Principal Lead Consultant in Imaging and Nuclear Medicine) was involved in the design of the study, delivery of the study and interpretation of the results.

**Lucy Pike** (<https://orcid.org/0000-0002-4381-3349>) (Clinical Scientist) was involved in the design of the study, and took responsibility for the PET accreditation and quality assurance, and chapter writing.

**Donald Sinclair** (Medical Physicist) was involved in PET accreditation and quality assurance aspects of the SPUtNik study, and chapter writing.

**Andrew Shah** (<https://orcid.org/0000-0001-8843-3406>) (Head of Radiation Protection) was involved in the design of the study and took responsibility for the DCE-CT accreditation and quality assurance, and chapter writing.

**Rosemary Eaton** (<https://orcid.org/0000-0001-6091-0477>) (Clinical Scientist) was involved in the design of the study and took responsibility for the DCE-CT accreditation and quality assurance, and chapter writing.

**Andrew Clegg** (<https://orcid.org/0000-0001-8938-7819>) (Professor of Health Services Research),  
**Valerio Benedetto** (<https://orcid.org/0000-0002-4683-0777>) (Research Associate, Health Economics)

and **James E Hill** (Senior Lecturer, Health Economics) were involved in the systematic review of cost-effectiveness, delivery of the study and chapter writing.

**Andrew Cook** (<https://orcid.org/0000-0002-6680-439X>) (Associate Director, SCTU) oversaw the study management and was involved in the interpretation of the results.

**Dimitrios Tzelis** (<https://orcid.org/0000-0002-6524-9048>) (Health Economist) and **Luke Vale** (<https://orcid.org/0000-0001-8574-8429>) (Professor of Health Economics) were involved in the health economics model development and analysis of model results, interpretation of the results and chapter writing.

**Lucy Brindle** (<https://orcid.org/0000-0002-8933-3754>) (Associate Professor in Early Diagnosis Research) developed and conducted the analysis for the IPCARD-SPN questionnaire substudy and was responsible for chapter writing.

**Jackie Madden** (<https://orcid.org/0000-0002-4604-731X>) (Trials Manager) and **Kelly Cozens** (<https://orcid.org/0000-0001-9592-9100>) (Senior Trials Manager) were responsible for study management.

**Louisa A Little** (<https://orcid.org/0000-0002-0083-7279>) (Senior Trials Manager) oversaw the study management, and was involved in the interpretation of the results and chapter writing.

**Kathrin Eichhorst** (<https://orcid.org/0000-0002-3666-4189>) (Data Co-ordinator) was responsible for data management.

**Patricia Moate** (Patient Representative) and **Chris McClement** (Patient Representative) were patient representatives on the Trial Management Group (TMG).

**Charles Peebles** (<https://orcid.org/0000-0003-3369-2173>) (Consultant Radiologist), **Anindo Banerjee** (Consultant in Thoracic Oncology and Tuberculosis), **Sai Han** (<https://orcid.org/0000-0001-5482-9648>) (Consultant in Nuclear Medicine), **Fat Wui Poon** (<https://orcid.org/0000-0003-1793-7580>) (Consultant Radiologist), **Ashley M Groves** (<https://orcid.org/0000-0003-0358-0795>) (Director, Institute of Nuclear Medicine), **Lutfi Kurban** (Consultant Radiologist), **Anthony J Frew** (Consultant Radiologist), **Matthew E Callister** (<https://orcid.org/0000-0001-8157-0803>) (Consultant in Respiratory Medicine), **Philip Crosbie** (<https://orcid.org/0000-0001-8941-4813>) (Clinical Senior Lecturer and Honorary Consultant in Respiratory Medicine), **Fergus V Gleeson** (<https://orcid.org/0000-0002-5121-3917>) (Professor of Radiology), **Kavitasagary Karunasaagarar** (<https://orcid.org/0000-0003-3151-4760>) (Radiology Consultant) and **Osei Kankam** (<https://orcid.org/0000-0003-3637-6133>) (Consultant Respiratory Physician) were involved in the design of the study and were responsible for recruiting participants.

**Steve George** (Consultant Clinical Epidemiologist and Co-Chief Investigator) was involved in the design of the study, delivery of the study and interpretation of the results.

All authors reviewed the final report.

## Publications

Qureshi NR, Rintoul RC, Miles KA, George S, Harris S, Madden J, *et al.* Accuracy and cost-effectiveness of dynamic contrast-enhanced CT in the characterisation of solitary pulmonary nodules – the SPUtNIK study. *BMJ Open Respir Res* 2016;**3**:e000156.

Qureshi NR, Shah A, Eaton RJ, Miles K, Gilbert FJ, Sputnik Investigators. Dynamic contrast enhanced CT in nodule characterization: how we review and report. *Cancer Imaging* 2016;**16**:16.

Weir-McCall JR, Joyce S, Clegg A, MacKay J, Baxter G, Dendl LM, *et al.* Dynamic contrast-enhanced computed tomography for the diagnosis of solitary pulmonary nodules: a systematic review and meta-analysis. *Eur Radiol* 2020;**30**:3310–23.

Gilbert FJ, Harris S, Miles KA, Weir-McCall JR, Qureshi NR, Campbell Rintoul R, *et al.* Comparative accuracy and cost-effectiveness of dynamic contrast-enhanced CT and positron emission tomography in the characterisation of solitary pulmonary nodules [published online ahead of print December 9 2021]. *Thorax* 2021.

### Data-sharing statement

Individual participant data will be made available, including data dictionaries, for approved data-sharing requests. Individual participant data will be shared that underlie the results reported in this article, after de-identification and normalisation of information (text, tables, figures and appendices). The study protocol and statistical analysis plan will also be available. Anonymous data will be available for request, from 3 months after publication of the article, to researchers who provide a completed data-sharing request form that describes a methodologically sound proposal, for the purpose of the approved proposal and, if appropriate, sign a data-sharing agreement. Data will be shared once all parties have signed relevant data-sharing documentation, covering SCTU conditions for sharing, and, if required, an additional data-sharing agreement from the sponsor. All data requests should be submitted to the corresponding author for consideration.

### Patient data

This work uses data provided by patients and collected by the NHS as part of their care and support. Using patient data is vital to improve health and care for everyone. There is huge potential to make better use of information from people's patient records, to understand more about disease, develop new treatments, monitor safety, and plan NHS services. Patient data should be kept safe and secure, to protect everyone's privacy, and it's important that there are safeguards to make sure that it is stored and used responsibly. Everyone should be able to find out about how patient data are used. #datasaveslives You can find out more about the background to this citation here: <https://understandingpatientdata.org.uk/data-citation>.

## References

1. Gilbert FJ, Harris S, Miles KA, Weir-McCall JR, Qureshi NR, Campbell Rintoul R, *et al*. Comparative accuracy and cost-effectiveness of dynamic contrast-enhanced CT and positron emission tomography in the characterisation of solitary pulmonary nodules [published online ahead of print December 9 2021]. *Thorax* 2021. <https://doi.org/10.1136/thoraxjnl-2021-216948>
2. Cancer Research UK. *Lung Cancer Incidence Statistics*. URL: [www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/lung-cancer/incidence](http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/lung-cancer/incidence) (accessed 25 September 2018).
3. Tanner NT, Dai L, Bade BC, Gebregziabher M, Silvestri GA. Assessing the generalizability of the National Lung Screening Trial: comparison of patients with stage 1 disease. *Am J Respir Crit Care Med* 2017;**196**:602–8. <https://doi.org/10.1164/rccm.201705-0914OC>
4. Barnett PG, Ananth L, Gould MK, Veterans Affairs Positron Emission Tomography Imaging in the Management of Patients with Solitary Pulmonary Nodules (VA SNAP) Cooperative Study Group. Cost and outcomes of patients with solitary pulmonary nodules managed with PET scans. *Chest* 2010;**137**:53–9. <https://doi.org/10.1378/chest.08-0529>
5. NHS. *NHS to Rollout Lung Cancer Scanning Trucks Across the Country*. URL: [www.england.nhs.uk/2019/02/lung-trucks/](http://www.england.nhs.uk/2019/02/lung-trucks/) (accessed 4 May 2019).
6. Black C, Bagust A, Boland A, Walker S, McLeod C, De Verteuil R, *et al*. The clinical effectiveness and cost-effectiveness of computed tomography screening for lung cancer: systematic reviews. *Health Technol Assess* 2006;**10**(3). <https://doi.org/10.3310/hta10030>
7. UK National Screening Committee. *UK NSC Recommendation on Lung Cancer Screening*. <https://legacyscreening.phe.org.uk/lungcancer> (accessed 31 December 2020).
8. Callister ME, Baldwin DR, Akram AR, Barnard S, Cane P, Draffan J, *et al*. British Thoracic Society guidelines for the investigation and management of pulmonary nodules. *Thorax* 2015;**70**(Suppl. 2):ii1–ii54. <https://doi.org/10.1136/thoraxjnl-2015-207168>
9. MacMahon H, Naidich DP, Goo JM, Lee KS, Leung ANC, Mayo JR, *et al*. Guidelines for management of incidental pulmonary nodules detected on CT images: from the Fleischner Society 2017. *Radiology* 2017;**284**:228–43. <https://doi.org/10.1148/radiol.2017161659>
10. MacMahon H, Austin JH, Gamsu G, Herold CJ, Jett JR, Naidich DP, *et al*. Guidelines for management of small pulmonary nodules detected on CT scans: a statement from the Fleischner Society. *Radiology* 2005;**237**:395–400. <https://doi.org/10.1148/radiol.2372041887>
11. National Institute for Health and Care Excellence (NICE). *Lung Cancer: Diagnosis and Management*. NICE guideline [NG122]. London: NICE; 2019.
12. Li ZZ, Huang YL, Song HJ, Wang YJ, Huang Y. The value of 18F-FDG-PET/CT in the diagnosis of solitary pulmonary nodules: a meta-analysis. *Medicine* 2018;**97**:e0130. <https://doi.org/10.1097/MD.00000000000010130>
13. Sim YT, Poon FW. Imaging of solitary pulmonary nodule – a clinical review. *Quant Imaging Med Surg* 2013;**3**:316–26.
14. Cronin P, Dwamena BA, Kelly AM, Carlos RC. Solitary pulmonary nodules: meta-analytic comparison of cross-sectional imaging modalities for diagnosis of malignancy. *Radiology* 2008;**246**:772–82. <https://doi.org/10.1148/radiol.2463062148>

15. Swensen SJ, Viggiano RW, Midthun DE, Müller NL, Sherrick A, Yamashita K, *et al.* Lung nodule enhancement at CT: multicenter study. *Radiology* 2000;**214**:73–80. <https://doi.org/10.1148/radiology.214.1.r00ja1473>
16. Field JK, Duffy SW, Baldwin DR, Whynes DK, Devaraj A, Brain KE, *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**:161–70. <https://doi.org/10.1136/thoraxjnl-2015-207140>
17. Yi CA, Lee KS, Kim BT, Choi JY, Kwon OJ, Kim H, *et al.* Tissue characterization of solitary pulmonary nodule: comparative study between helical dynamic CT and integrated PET/CT. *J Nucl Med* 2006;**47**:443–50.
18. Christensen JA, Nathan MA, Mullan BP, Hartman TE, Swensen SJ, Lowe VJ. Characterization of the solitary pulmonary nodule: 18F-FDG PET versus nodule-enhancement CT. *AJR Am J Roentgenol* 2006;**187**:1361–7. <https://doi.org/10.2214/AJR.05.1166>
19. Herder GJ, Van Tinteren H, Comans EF, Hoekstra OS, Teule GJ, Postmus PE, *et al.* Prospective use of serial questionnaires to evaluate the therapeutic efficacy of 18F-fluorodeoxyglucose (FDG) positron emission tomography (PET) in suspected lung cancer. *Thorax* 2003;**58**:47–51. <https://doi.org/10.1136/thorax.58.1.47>
20. Gould MK, Sanders GD, Barnett PG, Rydzak CE, Maclean CC, McClellan MB, Owens DK. Cost-effectiveness of alternative management strategies for patients with solitary pulmonary nodules. *Ann Intern Med* 2003;**138**:724–35. <https://doi.org/10.7326/0003-4819-138-9-200305060-00009>
21. Keith CJ, Miles KA, Griffiths MR, Wong D, Pitman AG, Hicks RJ. Solitary pulmonary nodules: accuracy and cost-effectiveness of sodium iodide FDG-PET using Australian data. *Eur J Nucl Med Mol Imaging* 2002;**29**:1016–23. <https://doi.org/10.1007/s00259-002-0833-2>
22. Gambhir SS, Shepherd JE, Shah BD, Hart E, Hoh CK, Valk PE, *et al.* Analytical decision model for the cost-effective management of solitary pulmonary nodules. *J Clin Oncol* 1998;**16**:2113–25. <https://doi.org/10.1200/JCO.1998.16.6.2113>
23. Shigeru K, Kiyoshi I, Masumi W, Hideo K, Shoichi K. Decision-tree sensitivity analysis for cost-effectiveness of chest 2-fluoro-2-D-[<sup>18</sup>F]fluorodeoxyglucose positron emission tomography in patients with pulmonary nodules (non-small-cell lung carcinoma) in Japan. *Chest* 2000;**117**:346–53. <https://doi.org/10.1378/chest.117.2.346>
24. Dietlein M, Weber K, Gandjour A, Moka D, Theissen P, Lauterbach KW, Schicha H. Cost-effectiveness of FDG-PET for the management of solitary pulmonary nodules: a decision analysis based on cost reimbursement in Germany. *Eur J Nucl Med* 2000;**27**:1441–56. <https://doi.org/10.1007/s002590000324>
25. Gugiatti A, Grimaldi A, Rossetti C, Lucignani G, De Marchis D, Borgonovi E, Fazio F. Economic analyses on the use of positron emission tomography for the work-up of solitary pulmonary nodules and for staging patients with non-small-cell-lung-cancer in Italy. *Q J Nucl Med Mol Imaging* 2004;**48**:49–61.
26. Comber LA, Keith CJ, Griffiths M, Miles KA. Solitary pulmonary nodules: impact of quantitative contrast-enhanced CT on the cost-effectiveness of FDG-PET. *Clin Radiol* 2003;**58**:706–11. [https://doi.org/10.1016/S0009-9260\(03\)00166-1](https://doi.org/10.1016/S0009-9260(03)00166-1)
27. Tsushima Y, Endo K. Analysis models to assess cost effectiveness of the four strategies for the work-up of solitary pulmonary nodules. *Med Sci Monit* 2004;**10**:MT65–MT72.

28. Lejeune C, Al Zahouri K, Woronoff-Lemsi MC, Arveux P, Bernard A, Binquet C, Guillemin F. Use of a decision analysis model to assess the medicoeconomic implications of FDG PET imaging in diagnosing a solitary pulmonary nodule. *Eur J Health Econ* 2005;**6**:203–14. <https://doi.org/10.1007/s10198-005-0279-0>
29. Pauls S, Buck AK, Halter G, Mottaghy FM, Mucic R, Bluemel C, *et al.* Performance of integrated FDG-PET/CT for differentiating benign and malignant lung lesions – results from a large prospective clinical trial. *Mol Imaging Biol* 2008;**10**:121–8. <https://doi.org/10.1007/s11307-007-0129-9>
30. Chang CY, Tzao C, Lee SC, Cheng CY, Liu CH, Huang WS, *et al.* Incremental value of integrated FDG-PET/CT in evaluating indeterminate solitary pulmonary nodule for malignancy. *Mol Imaging Biol* 2010;**12**:204–9. <https://doi.org/10.1007/s11307-009-0241-0>
31. Bisdas S, Spicer K, Rumboldt Z. Whole-tumor perfusion CT parameters and glucose metabolism measurements in head and neck squamous cell carcinomas: a pilot study using combined positron-emission tomography/CT imaging. *AJNR Am J Neuroradiol* 2008;**29**:1376–81. <https://doi.org/10.3174/ajnr.A1111>
32. Orlacchio A, Schillaci O, Antonelli L, D'Urso S, Sergiacomi G, Nicoli P, Simonetti G. Solitary pulmonary nodules: morphological and metabolic characterisation by FDG-PET-MDCT. *Radiol Med* 2007;**112**:157–73. <https://doi.org/10.1007/s11547-007-0132-x>
33. Ohno Y, Nishio M, Koyama H, Seki S, Tsubakimoto M, Fujisawa Y, *et al.* Solitary pulmonary nodules: comparison of dynamic first-pass contrast-enhanced perfusion area-detector CT, dynamic first-pass contrast-enhanced MR imaging, and FDG PET/CT. *Radiology* 2015;**274**:563–75. <https://doi.org/10.1148/radiol.14132289>
34. Ohno Y, Koyama H, Takenaka D, Nogami M, Maniwa Y, Nishimura Y, *et al.* Dynamic MRI, dynamic multidetector-row computed tomography (MDCT), and coregistered 2-[fluorine-18]-fluoro-2-deoxy-D-glucose-positron emission tomography (FDG-PET)/CT: comparative study of capability for management of pulmonary nodules. *J Magn Reson Imaging* 2008;**27**:1284–95. <https://doi.org/10.1002/jmri.21348>
35. Dabrowska M, Zukowska M, Krenke R, Domagała-Kulawik J, Maskey-Warzechowska M, Bogdan J, *et al.* Simplified method of dynamic contrast-enhanced computed tomography in the evaluation of indeterminate pulmonary nodules. *Respiration* 2010;**79**:91–6. <https://doi.org/10.1159/000213760>
36. Ohno Y, Koyama H, Matsumoto K, Onishi Y, Takenaka D, Fujisawa Y, *et al.* Differentiation of malignant and benign pulmonary nodules with quantitative first-pass 320-detector row perfusion CT versus FDG PET/CT. *Radiology* 2011;**258**:599–609. <https://doi.org/10.1148/radiol.10100245>
37. Ohno Y, Nishio M, Koyama H, Fujisawa Y, Yoshikawa T, Matsumoto S, Sugimura K. Comparison of quantitatively analyzed dynamic area-detector CT using various mathematic methods with FDG PET/CT in management of solitary pulmonary nodules. *AJR Am J Roentgenol* 2013;**200**:W593–602. <https://doi.org/10.2214/AJR.12.9197>
38. Marom E, Bruzzi J, Truong M. Extrathoracic PET/CT findings in thoracic malignancies. *J Thorac Imaging* 2006;**21**:154–66. <https://doi.org/10.1097/00005382-200605000-00007>
39. Qureshi NR, Rintoul RC, Miles KA, George S, Harris S, Madden J, *et al.* Accuracy and cost-effectiveness of dynamic contrast-enhanced CT in the characterisation of solitary pulmonary nodules-the SPUtNIk study. *BMJ Open Respir Res* 2016;**3**:e000156. <https://doi.org/10.1136/bmjresp-2016-000156>

40. National Institute for Health and Care Excellence. *Guide to the Methods of Technology Appraisal 2013*. URL: [www.nice.org.uk/process/pmg9/resources/guide-to-the-methods-of-technology-appraisal-2013-pdf-2007975843781](http://www.nice.org.uk/process/pmg9/resources/guide-to-the-methods-of-technology-appraisal-2013-pdf-2007975843781) (accessed 25 September 2018).
41. Gohagan JK, Marcus PM, Fagerstrom RM, Pinsky PF, Kramer BS, Prorok PC, *et al*. Final results of the Lung Screening Study, a randomized feasibility study of spiral CT versus chest X-ray screening for lung cancer. *Lung Cancer* 2005;**47**:9–15. <https://doi.org/10.1016/j.lungcan.2004.06.007>
42. Blanchon T, Bréchet JM, Grenier PA, Ferretti GR, Lemarié E, Milleron B, *et al*. Baseline results of the Depiscan study: a French randomized pilot trial of lung cancer screening comparing low dose CT scan (LDCT) and chest X-ray (CXR). *Lung Cancer* 2007;**58**:50–8. <https://doi.org/10.1016/j.lungcan.2007.05.009>
43. Alonzo TA, Pepe MS, Moskowitz CS. Sample size calculations for comparative studies of medical tests for detecting presence of disease. *Stat Med* 2002;**21**:835–53. <https://doi.org/10.1002/sim.1058>
44. Barrington SF, MacKewn JE, Schleyer P, Marsden PK, Mikhaeel NG, Qian W, *et al*. Establishment of a UK-wide network to facilitate the acquisition of quality assured FDG-PET data for clinical trials in lymphoma. *Ann Oncol* 2011;**22**:739–45. <https://doi.org/10.1093/annonc/mdq428>
45. Qureshi NR, Shah A, Eaton RJ, Miles K, Gilbert FJ, Sputnik investigators. Dynamic contrast enhanced CT in nodule characterization: how we review and report. *Cancer Imaging* 2016;**16**:16. <https://doi.org/10.1186/s40644-016-0074-4>
46. McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM, Clifford T, *et al*. Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies: the PRISMA-DTA statement. *JAMA* 2018;**319**:388–96. <https://doi.org/10.1001/jama.2017.19163>
47. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, *et al*. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;**155**:529–36. <https://doi.org/10.7326/0003-4819-155-8-201110180-00009>
48. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *J Stat Softw* 2015;**67**:48. <https://doi.org/10.18637/jss.v067.i01>
49. Harbord RM, Egger M, Sterne JAC. A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. *Stat Med* 2006;**25**:3443–57. <https://doi.org/10.1002/sim.2380>
50. Basso Dias A, Zanon M, Altmayer S, Sartori Pacini G, Henz Concatto N, Watte G, *et al*. Fluorine 18-FDG PET/CT and diffusion-weighted MRI for malignant versus benign pulmonary lesions: a meta-analysis. *Radiology* 2019;**290**:525–34. <https://doi.org/10.1148/radiol.2018181159>
51. Ducharme J, Goertzen AL, Patterson J, Demeter S. Practical aspects of 18F-FDG PET when receiving 18F-FDG from a distant supplier. *J Nucl Med Technol* 2009;**37**:164–9. <https://doi.org/10.2967/jnmt.109.062950>
52. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. *Methods for the Economic Evaluation of Health Care Programmes*. 4th edn. Oxford: Oxford University Press; 2015.
53. Philips Z, Ginnelly L, Sculpher M, Claxton K, Golder S, Riemsma R, *et al*. Review of guidelines for good practice in decision-analytic modelling in health technology assessment. *Health Technol Assess* 2004;**8**(36). <https://doi.org/10.3310/hta8360>
54. Higgins JPT, Green S. *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.1.0. Updated March 2011. London: The Cochrane Collaboration; 2011.

55. Deppen SA, Davis WT, Green EA, Rickman O, Aldrich MC, Fletcher S, *et al.* Cost-effectiveness of initial diagnostic strategies for pulmonary nodules presenting to thoracic surgeons. *Ann Thorac Surg* 2014;**98**:1214–22. <https://doi.org/10.1016/j.athoracsur.2014.05.025>
56. Reitsma JB, Rutjes AWS, Whiting P, Vlassov VV, Leeflang MMG, Deeks JJ. Chapter 9: Assessing Methodological Quality. In Deeks JJ, Bossuyt PM, Gatsonis C, editors. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*. Version 1.0.0. London: The Cochrane Collaboration, 2009. URL: <http://srdta.cochrane.org/> (accessed 25 September 2018).
57. Cummings S, Lillington G, Richard R. Managing solitary pulmonary nodules. *Am Rev Respir Dis* 1986;**134**:453–60.
58. Steele JD, Buell P. Asymptomatic solitary pulmonary nodules. Host survival, tumor size, and growth rate. *J Thorac Cardiovasc Surg* 1973;**65**:140–51. [https://doi.org/10.1016/S0022-5223\(19\)40835-0](https://doi.org/10.1016/S0022-5223(19)40835-0)
59. Beck JR, Kassirer JP, Pauker SG. A convenient approximation of life expectancy (the 'DEALE'). I. Validation of the method. *Am J Med* 1982;**73**:883–8. [https://doi.org/10.1016/0002-9343\(82\)90786-0](https://doi.org/10.1016/0002-9343(82)90786-0)
60. Beck JR, Pauker SG, Gottlieb JE, Klein K, Kassirer JP. A convenient approximation of life expectancy (the 'DEALE'). II. Use in medical decision-making. *Am J Med* 1982;**73**:889–97. [https://doi.org/10.1016/0002-9343\(82\)90787-2](https://doi.org/10.1016/0002-9343(82)90787-2)
61. Ohno Y, Hatabu H, Takenaka D, Higashino T, Watanabe H, Ohbayashi C, Sugimara K. CT-guided transthoracic needle aspiration biopsy of small ( $\leq 20$  mm) solitary pulmonary nodules. *Am J Roentgenol* 2003;**180**:1665–9. <https://doi.org/10.2214/ajr.180.6.1801665>
62. Doubilet P, Begg CB, Weinstein MC, Braun P, McNeil BJ. Probabilistic sensitivity analysis using Monte Carlo Simulation. A practical approach. *Med Decis Making* 1985;**5**:157–77. <https://doi.org/10.1177/0272989X8500500205>
63. Anderson R. Systematic reviews of economic evaluations: utility or futility? *Health Econ* 2010;**19**:350–64. <https://doi.org/10.1002/hec.1486>
64. Kielhorn A, Graf von der Schulenburg JM. *The Health Economics Handbook*. Auckland: Adis International Ltd; 2000.
65. Department of Health and Social Care. *NHS Reference Costs 2017/18*. URL: <https://improvement.nhs.uk/resources/reference-costs/#rc1718> (accessed 25 September 2018).
66. Han Y, Kim HJ, Kong KA, Kim SJ, Lee SH, Ryu YJ, *et al.* Diagnosis of small pulmonary lesions by transbronchial lung biopsy with radial endobronchial ultrasound and virtual bronchoscopic navigation versus CT-guided transthoracic needle biopsy: a systematic review and meta-analysis. *PLOS One* 2018;**13**:e0191590. <https://doi.org/10.1371/journal.pone.0191590>
67. Wiener RS, Schwartz LM, Woloshin S, Welch HG. Population-based risk for complications after transthoracic needle lung biopsy of a pulmonary nodule: an analysis of discharge records. *Ann Intern Med* 2011;**155**:137–44. <https://doi.org/10.7326/0003-4819-155-3-201108020-00003>
68. Zhao J, Yorke ED, Li L, Kavanagh BD, Li XA, Das S, *et al.* Simple factors associated with radiation-induced lung toxicity after stereotactic body radiation therapy of the thorax: a pooled analysis of 88 studies. *Int J Radiat Oncol Biol Phys* 2016;**95**:1357–66. <https://doi.org/10.1016/j.ijrobp.2016.03.024>
69. National Institute for Health and Care Excellence (NICE) Guideline Updates Team. *Lung Cancer Update: Evidence Reviews for the Clinical and Cost Effectiveness of Different Radiotherapy Regimens with Curative Intent for NSCLC*. NICE Guideline NG122: Evidence Reviews. London: NICE; 2019.

70. London Cancer Alliance. *LCA Acute Oncology Clinical Guidelines*. London: London Cancer Alliance; 2014.
71. Joint Formulary Committee. *British National Formulary* (online). London: BMJ Group and Pharmaceutical Press. URL: [www.medicinescomplete.com](http://www.medicinescomplete.com) (accessed 15 July 2019).
72. Paix A, Noel G, Falcoz PE, Levy P. Cost-effectiveness analysis of stereotactic body radiotherapy and surgery for medically operable early stage non small cell lung cancer. *Radiother Oncol* 2018;**128**:534–40. <https://doi.org/10.1016/j.radonc.2018.04.013>
73. Echevarria C, Gray J, Hartley T, Steer J, Miller J, Simpson AJ, *et al*. Home treatment of COPD exacerbation selected by DECAF score: a non-inferiority, randomised controlled trial and economic evaluation. *Thorax* 2018;**73**:713–22. <https://doi.org/10.1136/thoraxjnl-2017-211197>
74. Curtis L, Burns A. *Unit Costs of Health and Social Care 2018*. Canterbury: Personal Social Services Research Unit, University of Kent; 2018.
75. Accordino MK, Wright JD, Buono D, Neugut AI, Hershman DL. Trends in use and safety of image-guided transthoracic needle biopsies in patients with cancer. *J Oncol Pract* 2015;**11**:e351–9. <https://doi.org/10.1200/JOP.2014.001891>
76. Ben-Aharon O, Magnezi R, Leshno M, Goldstein DA. Median survival or mean survival: which measure is the most appropriate for patients, physicians, and policymakers? *Oncologist* 2019;**24**:1469–78. <https://doi.org/10.1634/theoncologist.2019-0175>
77. Office for National Statistics. *National Life Tables*. Newport: Office for National Statistics; 2018.
78. Khakwani A, *et al*. P2.16-21 Post-treatment survival difference between lobectomy and stereotactic ablative radiotherapy in stage 1 non-small cell lung cancer in England. *J Thorac Oncol* 2018;**13**:S839. <https://doi.org/10.1016/j.jtho.2018.08.1496>
79. Rosen JE, Keshava HB, Yao X, Kim AW, Detterbeck FC, Boffa DJ. The natural history of operable non-small cell lung cancer in the National Cancer Database. *Ann Thorac Surg* 2016;**101**:1850–5. <https://doi.org/10.1016/j.athoracsur.2016.01.077>
80. Kind P, Hardman G, Macran S. *UK Population Norms for EQ-5D*. Centre for Health Economics discussion paper 172. York: Centre for Health Economics, University of York; 1999.
81. Doyle S, Lloyd A, Walker M. Health state utility scores in advanced non-small cell lung cancer. *Lung Cancer* 2008;**62**:374–80. <https://doi.org/10.1016/j.lungcan.2008.03.019>
82. Brindle L, Dowswell G, James EP, Clifford S, Ocansey L, Hamilton W, Banerjee A. *Using a Participant-completed Questionnaire to Identify Symptoms that Predict Chest and Respiratory Disease (IPCARD): A Feasibility Study*. 2015. URL: <https://eprints.soton.ac.uk/378960/> (accessed 25 September 2018).
83. Brindle L, Pope C, Corner J, Leydon G, Banerjee A. Eliciting symptoms interpreted as normal by patients with early-stage lung cancer: could GP elicitation of normalised symptoms reduce delay in diagnosis? Cross-sectional interview study. *BMJ Open* 2012;**2**:e001977. <https://doi.org/10.1136/bmjopen-2012-001977>
84. Young RP, Hopkins RJ, Christmas T, Black PN, Metcalf P, Gamble GD. COPD prevalence is increased in lung cancer, independent of age, sex and smoking history. *Eur Respir J* 2009;**34**:380–6. <https://doi.org/10.1183/09031936.00144208>
85. Hamilton W, Peters TJ, Round A, Sharp D. What are the clinical features of lung cancer before the diagnosis is made? A population based case-control study. *Thorax* 2005;**60**:1059–65. <https://doi.org/10.1136/thx.2005.045880>

86. Cassidy A, Myles JP, van Tongeren M, Page RD, Liloglou T, Duffy SW, Field JK. The LLP risk model: an individual risk prediction model for lung cancer. *Br J Cancer* 2008;**98**:270–6. <https://doi.org/10.1038/sj.bjc.6604158>
87. Hamilton W. The CAPER studies: five case-control studies aimed at identifying and quantifying the risk of cancer in symptomatic primary care patients. *Br J Cancer* 2009;**101**(Suppl. 2):80–6. <https://doi.org/10.1038/sj.bjc.6605396>
88. Kadota K, Colovos C, Suzuki K, Rizk NP, Dunphy MP, Zabor EC, *et al.* FDG-PET SUVmax combined with IASLC/ATS/ERS histologic classification improves the prognostic stratification of patients with stage I lung adenocarcinoma. *Ann Surg Oncol* 2012;**19**:3598–605. <https://doi.org/10.1245/s10434-012-2414-3>
89. Lee KS, Yi C, Jeong SY, Jeong YJ, Kim S, Chung MJ, *et al.* Solid or partly solid solitary pulmonary nodules: their characterization using contrast wash-in and morphologic features at helical CT. *Chest* 2007;**131**:1516–25. <https://doi.org/10.1378/chest.06-2526>
90. Swensen SJ, Morin RL, Schueler BA, Brown LR, Cortese DA, Pairolero PC, Brutinel WM. Solitary pulmonary nodule: CT evaluation of enhancement with iodinated contrast material – a preliminary report. *Radiology* 1992;**182**:343–7. <https://doi.org/10.1148/radiology.182.2.1732947>
91. The National Lung Screening Trial Research, Team. Results of initial low-dose computed tomographic screening for lung cancer. *New Engl J Med* 2013;**368**:1980–91. <https://doi.org/10.1056/NEJMoa1209120>
92. Horeweg N, van Rosmalen J, Heuvelmans MA, van der Aalst CM, Vliegenthart R, Scholten ET, *et al.* Lung cancer probability in patients with CT-detected pulmonary nodules: a prespecified analysis of data from the NELSON trial of low-dose CT screening. *Lancet Oncol* 2014;**15**:1332–41. [https://doi.org/10.1016/S1470-2045\(14\)70389-4](https://doi.org/10.1016/S1470-2045(14)70389-4)
93. Leeftang MMG, Rutjes AWS, Reitsma JB, Hooft L, Bossuyt PMM. Variation of a test's sensitivity and specificity with disease prevalence. *Can Med Assoc J* 2013;**185**:E537–44. <https://doi.org/10.1503/cmaj.121286>
94. Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, *et al.* Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011;**365**:395–409. <https://doi.org/10.1056/NEJMoa1102873>
95. Goodman A. *NELSON Trial: 'Call to Action' for Lung Cancer CT Scanning of High-risk Individuals.* The ASCO Post; 2018. URL: [www.ascopost.com/issues/october-25-2018/nelson-trial/](http://www.ascopost.com/issues/october-25-2018/nelson-trial/) (accessed 4 May 2019).
96. 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**:1732–42. <https://doi.org/10.1016/j.jtho.2019.05.044>
97. De Koning H, Van Der Aalst C, Ten Haaf K, Oudkerk M. PL02.05 Effects of volume CT lung cancer screening: mortality results of the NELSON randomised-controlled population based trial. *J Thorac Oncol* 2018;**13**:S185. <https://doi.org/10.1016/j.jtho.2018.08.012>
98. Crosbie PA, Balata H, Evison M, Atack M, Bayliss-Brideaux V, Colligan D, *et al.* Implementing lung cancer screening: baseline results from a community-based 'Lung Health Check' pilot in deprived areas of Manchester. *Thorax* 2019;**74**:405–9. <https://doi.org/10.1136/thoraxjnl-2017-211377>
99. Swensen SJ, Brown LR, Colby TV, Weaver AL. Pulmonary nodules: CT evaluation of enhancement with iodinated contrast material. *Radiology* 1995;**194**:393–8. <https://doi.org/10.1148/radiology.194.2.7824716>

100. Yamashita K, Matsunobe S, Tsuda T, Nemoto T, Matsumoto K, Miki H, Konishi J. Solitary pulmonary nodule: preliminary study of evaluation with incremental dynamic CT. *Radiology* 1995;**194**:399–405. <https://doi.org/10.1148/radiology.194.2.7824717>
101. Swensen SJ, Brown LR, Colby TV, Weaver AL, Midthun DE. Lung nodule enhancement at CT: prospective findings. *Radiology* 1996;**201**:447–55. <https://doi.org/10.1148/radiology.201.2.8888239>
102. Potente G, Iacari V, Caimi M. The challenge of solitary pulmonary nodules: HRCT evaluation. *Comput Med Imaging Graph* 1997;**21**:39–46. [https://doi.org/10.1016/S0895-6111\(96\)00071-7](https://doi.org/10.1016/S0895-6111(96)00071-7)
103. Zhang M, Kono M. Solitary pulmonary nodules: evaluation of blood flow patterns with dynamic CT. *Radiology* 1997;**205**:471–8. <https://doi.org/10.1148/radiology.205.2.9356631>
104. Kim JH, Kim HJ, Lee KH, Kim KH, Lee HL. Solitary pulmonary nodules: a comparative study evaluated with contrast-enhanced dynamic MR imaging and CT. *J Comput Assist Tomogr* 2004;**28**:766–75. <https://doi.org/10.1097/00004728-200411000-00007>
105. Choi EJ, Jin GY, Han YM, Lee YS, Kweon KS. Solitary pulmonary nodule on helical dynamic CT scans: analysis of the enhancement patterns using a computer-aided diagnosis (CAD) system. *Korean J Radiol* 2008;**9**:401–8. <https://doi.org/10.3348/kjr.2008.9.5.401>
106. Bayraktaroglu S, Savaş R, Basoglu OK, Cakan A, Mogulkoc N, Cagirici U, Alper H. Dynamic computed tomography in solitary pulmonary nodules. *J Comput Assist Tomogr* 2008;**32**:222–7. <https://doi.org/10.1097/RCT.0b013e318136e29d>
107. Bai RJ, Cheng XG, Qu H, Shen BZ, Han MJ, Wu ZH. Solitary pulmonary nodules: comparison of multi-slice computed tomography perfusion study with vascular endothelial growth factor and microvessel density. *Chin Med J* 2009;**122**:541–7.
108. Jiang NC, Han P, Zhou CK, Zheng JL, Shi HS, Xiao J. Dynamic enhancement patterns of solitary pulmonary nodules at multi-detector row CT and correlation with vascular endothelial growth factor and microvessel density. *Ai Zheng* 2009;**28**:164–9.
109. Li Y, Yang ZG, Chen TW, Yu JQ, Sun JY, Chen HJ. First-pass perfusion imaging of solitary pulmonary nodules with 64-detector row CT: comparison of perfusion parameters of malignant and benign lesions. *Br J Radiol* 2010;**83**:785–90. <https://doi.org/10.1259/bjr/58020866>
110. Shu SJ, Liu BL, Jiang HJ. Optimization of the scanning technique and diagnosis of pulmonary nodules with first-pass 64-detector-row perfusion VCT. *Clin Imaging* 2013;**37**:256–64. <https://doi.org/10.1016/j.clinimag.2012.05.004>
111. Ribeiro SM, Ruiz RL, Yoo HH, Cataneo DC, Cataneo AJ. Proposal to utilize simplified Swensen protocol in diagnosis of isolated pulmonary nodule. *Acta Radiol* 2013;**54**:757–64. <https://doi.org/10.1177/0284185113481695>
112. Ye XD, Ye JD, Yuan Z, Li WT, Xiao XS. Dynamic CT of solitary pulmonary nodules: comparison of contrast medium distribution characteristic of malignant and benign lesions. *Clin Transl Oncol* 2014;**16**:49–56. <https://doi.org/10.1007/s12094-013-1039-8>
113. Baldwin DR, Eaton T, Kolbe J, Christmas T, Milne D, Mercer J, et al. Management of solitary pulmonary nodules: how do thoracic computed tomography and guided fine needle biopsy influence clinical decisions? *Thorax* 2002;**57**:817–22. <https://doi.org/10.1136/thorax.57.9.817>
114. Gupta S, Krishnamurthy S, Broemeling LD, Morello FA Jr, Wallace MJ, Ahrar K, et al. Small ( $\leq 2$ -cm) subpleural pulmonary lesions: short-versus long-needle-path CT-guided biopsy—comparison of diagnostic yields and complications. *Radiol* 2005;**234**:631–7. <https://doi.org/10.1148/radiol.2342031423>

115. Hayashi N, Sakai T, Kitagawa M, Kimoto T, Inagaki R, Ishii Y, Noriki S, Imamura Y. CT-guided biopsy of pulmonary nodules less than 3 cm: usefulness of the spring-operated core biopsy needle and frozen-section pathologic diagnosis. *AJR Am J Roentgenol* 1998;**170**:329–31. <https://doi.org/10.2214/ajr.170.2.9456939>
116. Jin KN, Park CM, Goo JM, Lee HJ, Lee Y, Kim JI, et al. Initial experience of percutaneous transthoracic needle biopsy of lung nodules using C-arm cone-beam CT systems. *Eur Radiol* 2010;**20**:2108–15. <https://doi.org/10.1007/s00330-010-1783-x>
117. Ohno Y, Hatabu H, Takenaka D, Imai M, Ohbatashi C, Sugimura K. Transthoracic CT-guided biopsy with multiplanar reconstruction image improves diagnostic accuracy of solitary pulmonary nodules. *Eur J Radiol* 2004;**51**:160–8. [https://doi.org/10.1016/S0720-048X\(03\)00216-X](https://doi.org/10.1016/S0720-048X(03)00216-X)
118. Romano M, Griffo S, Gentile M, Mainenti PP, Tamburrini O, Iaccarino V, et al. CT guided percutaneous fine needle biopsy of small lung lesions in outpatients. Safety and efficacy of the procedure compared to inpatients. *Radiol Med* 2004;**108**:275–82.
119. Santambrogio L, Nosotti M, Bellaviti N, Pavoni G, Radice F, Caputo V. CT-guided fine-needle aspiration cytology of solitary pulmonary nodules: a prospective, randomized study of immediate cytologic evaluation. *Chest* 1997;**112**:423–5. <https://doi.org/10.1378/chest.112.2.423>
120. Tsukada H, Satou T, Iwashima A, Souma T. Diagnostic accuracy of CT-guided automated needle biopsy of lung nodules. *AJR Am J Roentgenol* 2000;**175**:239–43. <https://doi.org/10.2214/ajr.175.1.1750239>
121. Wagnetz U, Menezes RJ, Boerner S, Paul NS, Wagnetz D, Keshavjee S, et al. CT screening for lung cancer: implication of lung biopsy recommendations. *AJR Am J Roentgenol* 2012;**198**:351–8. <https://doi.org/10.2214/ajr.11.6726>
122. Westcott JL, Rao N, Colley DP. Transthoracic needle biopsy of small pulmonary nodules. *Radiol* 1997;**202**:97–103. <https://doi.org/10.1148/radiology.202.1.8988197>
123. Wang W, Yu L, Wang Y, Zhang Q, Chi C, Zhan P, Xu C. Radial EBUS versus CT-guided needle biopsy for evaluation of solitary pulmonary nodules. *Oncotarget* 2018;**9**:15122–31. <https://doi.org/10.18632/oncotarget.23952>
124. Xu C, Yuan Q, Chi C, Zhang Q, Wang Y, Wang W, et al. Computed tomography-guided percutaneous transthoracic needle biopsy for solitary pulmonary nodules in diameter less than 20 mm. *Medicine* 2018;**97**:e0154. <https://doi.org/10.1097/MD.00000000000010154>
125. Liu M, Huang J, Xu Y, He X, Li L, Lü Y, et al. MR-guided percutaneous biopsy of solitary pulmonary lesions using a 1.0-T open high-field MRI scanner with respiratory gating. *Eur Radiol* 2017;**27**:1459–66. <https://doi.org/10.1007/s00330-016-4518-9>
126. Liu S, Li C, Yu X, Liu M, Fan T, Chen D, et al. Diagnostic accuracy of MRI-guided percutaneous transthoracic needle biopsy of solitary pulmonary nodules. *Cardiovasc Intervent Radiol* 2015;**38**:416–21. <https://doi.org/10.1007/s00270-014-0915-0>
127. Sa YJ, Kim JJ, Kim YD, Sim SB, Moon SW. A new protocol for concomitant needle aspiration biopsy and localization of solitary pulmonary nodules. *J Cardiothorac Surg* 2015;**10**:104. <https://doi.org/10.1186/s13019-015-0312-z>
128. Yang W, Sun W, Li Q, Yao Y, Lv T, Zeng J, et al. Diagnostic accuracy of CT-guided transthoracic needle biopsy for solitary pulmonary nodules. *PLOS ONE* 2015;**10**:e0131373. <https://doi.org/10.1371/journal.pone.0131373>
129. Lee KJ, Han YM, Jin GY, Song JS. Predicting factors for conversion from fluoroscopy guided percutaneous transthoracic needle biopsy to cone-beam CT guided percutaneous transthoracic needle biopsy. *J Korean Soc Radiol* 2015;**73**:216–24. <https://doi.org/10.3348/jksr.2015.73.4.216>

130. Kaaki S, Kldane B, Srinathan S, Tan L, Buduhan G. Is tissue still the issue? Lobectomy for suspicious lung nodules without confirmation of malignancy. *J Surg Oncol* 2018;**117**:977–84. <https://doi.org/10.1002/jso.25003>
131. Lu R, Mei J, Zhao D, Jiang Z, Xiao H, Wang M, Ma N. Concomitant thoracoscopic surgery for solitary pulmonary nodule and atrial fibrillation. *Interact Cardiovasc Thorac Surg* 2018;**26**:402–6. <https://doi.org/10.1093/icvts/ivx346>
132. Yang SM, Wang ML, Hung MH, Hsu HH, Cheng YJ, Chen JS. Tubeless uniportal thoracoscopic wedge resection for peripheral lung nodules. *Ann Thorac Surg* 2017;**103**:462–8. <https://doi.org/10.1016/j.athoracsur.2016.09.006>
133. Li S, Jiang L, Ang KL, Chen H, Dong Q, Yang H, *et al.* New tubeless video-assisted thoracoscopic surgery for small pulmonary nodules. *Eur J Cardiothorac Surg* 2017;**51**:689–93.
134. Qi H, Wan C, Zhang L, Wang J, Song Z, Zhang R, *et al.* Early effective treatment of small pulmonary nodules with video-assisted thoracoscopic surgery combined with CT-guided dual-barbed hookwire localization. *Oncotarget* 2017;**8**:38793–801. <https://doi.org/10.18632/oncotarget.17044>
135. Müller J, Putora PM, Scheider T, Zeisel C, Brutsche M, *et al.* Handheld Single Photon Emission Computed Tomography (handheld SPECT) navigated video-assisted thoracoscopic surgery of computer tomography-guided radioactively marked pulmonary lesions. *Interact Cardiovasc Thorac Surg* 2016;**23**:345–50. <https://doi.org/10.1093/icvts/ivw136>
136. Ghaly G, Kamel M, Nasar A, Paul S, Lee PC, Port JL, *et al.* Video-assisted thoracoscopic surgery Is a safe and effective alternative to thoracotomy for anatomical segmentectomy in patients with clinical stage I non-small cell lung cancer. *Ann Thorac Surg* 2016;**101**:465–72. <https://doi.org/10.1016/j.athoracsur.2015.06.112>
137. Gill RR, Zheng Y, Barlow JS, Jayender J, Girard EE, Hartigan PM, *et al.* Image-guided video assisted thoracoscopic surgery (iVATS) – phase I–II clinical trial. *J Surg Oncol* 2015;**112**:18–25. <https://doi.org/10.1002/jso.23941>
138. Ren M, Meng Q, Zhou W, Kong F, Yang B, Yuan J, *et al.* Comparison of short-term effect of thoracoscopic segmentectomy and thoracoscopic lobectomy for the solitary pulmonary nodule and early-stage lung cancer. *Onco Targets Ther* 2014;**7**:1343–7. <https://doi.org/10.2147/OTT.S62880>
139. Cho J, Ko SJ, Kim SJ, Lee YJ, Park JS, Cho YJ, *et al.* Surgical resection of nodular ground-glass opacities without percutaneous needle aspiration or biopsy. *BMC Cancer* 2014;**14**:838. <https://doi.org/10.1186/1471-2407-14-838>

# Appendix 1 Patient and public involvement in the SPUtNik study

Throughout the study, there was patient and public involvement (PPI) representation on both the TMG and the Trial Steering Committee (TSC).

## Trial Management Group

Chris McClement was the original PPI member of the trial development team and was fully engaged and involved in the development of the final study protocol. Sadly, before the trial opened to recruitment, Chris McClement passed away. Patricia Moate took over as PPI representative on the TMG and assisted the trial team thereafter, and was an integral part of the TMG.

Patricia Moate was a dedicated PPI member and was involved in most of the TMG meetings prior to becoming ill in 2018. She was fully committed at all times and added valuable comments and points for discussion at all meetings that she attended. Although disabled through her condition, Patricia still managed to attend meetings in person, join conference calls or submit, via e-mail, comments on documents for consideration by the team.

Patricia made the following contributions:

- attended TMG face-to-face meeting when they took place and she was able to do so
- attended TMG conference calls and reviewed all necessary documents that were tabled for discussion at each conference
- reviewed the protocol when applicable and contributed to amendments
- reviewed all patient-facing documentation, providing valuable advice and points for discussion
- commented on patient pathways and gave advice regarding burden on the patient to assist with troubleshooting patient recruitment issues in the earlier, and also the later, phases of the trial.

Some specific contributions:

- Recruitment logs – Patricia was involved in all discussions regarding recruitment issues among sites and provided some valuable input. For example, she raised some concern about a site that was recording ‘panic attack’ as the main reason consented patients had not undergone DCE-CT, whereas other sites had not. This was investigated by the trial team at the time, and it was discovered that differences in language/terminology used between sites was the problem. This led the trial team to ask sites to be more specific in terms of language on recruitment logs to enable us to provide a more accurate CONSORT flow diagram.
- Recruitment difficulties and delays – certain legalities concerning PET and DCE-CT being performed at an alternative trust for one particular site led to a delay of > 1 year in opening to recruitment. The site eventually opened, but, because of these legalities, was soon closed again. Patricia was very concerned that patients were being prevented from taking part in valuable research because R&D departments between collaborating NHS departments could not solve this problem. She advised that we write to the head of the trust to highlight this as an issue. On further discussion within the TMG, it was decided that a letter be sent to the head of the trust and the local CRN.
- Substudy – it was decided that the IPCARD-SPN questionnaires be sent a second time to SPUtNik patients at 18 months’ study participation. This was widely discussed within the TMG, as it raised questions about who the questionnaires should and should not be sent to and how we would ensure that it was not sent to deceased or ill patients. Patricia contributed to these discussions and helped the team rationalise their decision to re-send the questionnaire and helped in the development of the associated letters of information and patient information documentation.

## **Trial Steering Committee**

Lisa Lamond was recruited to the TSC at its inception. Lisa made the following contributions:

- attended TSC face-to-face meetings when they took place
- attended TSC conference calls and reviewed all necessary documents that were tabled for discussion at each conference
- reviewed the TSC reports and actively contributed to discussions regarding trial progress and any issues tabled for discussion
- attended the final TSC results meeting and reviewed the lay summary for the final report, and will support the team in disseminating the results.

## Appendix 2 Methods tables

TABLE 27 List of recruiting sites and PET centres

Recruiting site	PET centre	Scanners
Aberdeen	Aberdeen	GE Discovery STE (GE Healthcare) GE Discovery 710 (GE Healthcare)
Brighton, Hastings and Worthing	Brighton	Siemens Biograph 64 (Siemens Healthineers AG)
Edinburgh	Edinburgh	Siemens Biograph 128 (Siemens Healthineers AG)
Glasgow	Glasgow	GE Discovery STE GE Discovery 690 (GE Healthcare)
Leicester	Mobile <sup>a</sup>	1 × Siemens Biograph 6 (Siemens Healthineers AG) 2 × GE Discovery 710
Leeds	Leeds	GE Discovery 690
Manchester	Christie Central Manchester Mobile <sup>a</sup>	GE Discovery STE Siemens Biograph mCT (Siemens Healthineers AG) GE Discovery 710
Nottingham	Nottingham	GE Discovery 710
Oxford	Oxford	GE Discovery 690
Royal Papworth Hospital	Mobile <sup>a</sup> Addenbrooke's Hospital	4 × Siemens Biograph 6 GE Discovery 690
Southampton	Mobile <sup>a</sup> Portsmouth	2 × Siemens Biograph 6 Siemens Biograph mCT
UCLH	UCLH Royal Free Hospital	GE Discovery VCT (GE Healthcare) Siemens Biograph mCT
Cheltenham, Weston Park, Worcester	Cheltenham	Philips Gemini GXL (Koninklijke Philips N.V., Amsterdam, the Netherlands) Siemens Biograph 128

UCLH, University College London Hospital.

<sup>a</sup> Eight mobile scanners were used across the sites, with the same PET/CT scanner used at more than one recruiting site.

TABLE 28 List of scanners used by site

Recruiting site	DCE-CT centre	Scanners
Aberdeen	Aberdeen	GE Discovery 750HD (GE Healthcare)
Brighton, Hastings and Worthing	Brighton	Siemens Biograph 64
Edinburgh	Edinburgh	Siemens Biograph 128
Glasgow	Glasgow	GE Discovery 690
Leicester	Leicester	Siemens Definition Flash (Siemens Healthineers AG)
Leeds	Leeds	GE Discovery 690
Manchester	Wythenshawe	Siemens Somatom Sensation 16 (Siemens Healthineers AG)
Nottingham	Nottingham	Siemens Somatom Definition AS+ (Siemens Healthineers AG)
Oxford	Oxford	GE Discovery 690
Royal Papworth Hospital	Royal Papworth Hospital I <sup>a</sup>	Siemens Definition (Siemens Healthineers AG)
	Royal Papworth Hospital II	Siemens Force (Siemens Healthineers AG)
	Royal Papworth Hospital III	Siemens Definition AS (Siemens Healthineers AG)
Southampton	Southampton	Siemens Sensation 64 (Siemens Healthineers AG)
UCLH	UCLH	GE Discovery VCT
Weston General Hospital	Weston-super-Mare	Siemens Somatom Definition AS+
Worcester	Worcester I <sup>b</sup>	Toshiba Aquilion 64 (CT2) (Toshiba Corporation, Tokyo, Japan)
	Worcester II	Toshiba Aquilion CX 64 (CT1) (Toshiba Corporation)

UCLH, University College London Hospital.

a Scanner decommissioned for clinical work during the course of the trial. Second and third scanners accredited.

b Scanner decommissioned for research work before any SPUtNik patient scanned. Second scanner accredited.

## Appendix 3 The SPUtNik trial: radiographer quality control measurements

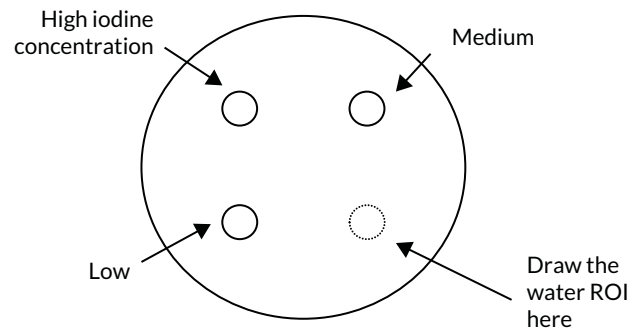
These QC measurements are designed to monitor variation in the CT scanner performance. Patient diagnosis in this trial is dependent on the accuracy of the CT HU measured during the DCE-CT procedure. When not in use, the trial phantom should be stored securely at room temperature.

These measurements should be carried out before each trial participant is scanned:

1. Carry out an air calibration at the start of the day and carry out routine manufacturer QC tests.
2. Before scanning, shake the calibration phantom gently to ensure that the iodine inserts are well mixed.
3. Tip the phantom to position any large air bubbles behind the bubble traps.
4. Align the phantom with the plugs at the 'head' end of the couch, approximately 5 cm from the internal axial laser. Line up the phantom marks with the sagittal and coronal lasers.
5. Load up the SPUtNik QA protocol and carry out a topogram of the phantom.
6. Position the scan range over the centre of the phantom and scan.
7. Using circular ROIs of area 3 cm<sup>2</sup> (Siemens, Toshiba) or 300 mm<sup>2</sup> (GE), measure CT numbers on the central slice or the closest slice containing no bubbles.
8. Measure inside the three circular regions (high, medium and low iodine concentrations), and also in the background water.
9. Record values in the trial QA folder or spreadsheet –
  - If CT numbers are outside tolerance, check the phantom alignment and repeat the QC tests.
  - If CT numbers are still outside tolerance, repeat the air calibration and then repeat QC tests.
  - If CT numbers are still outside tolerance, contact the Medical Physics Department at Mount Vernon Hospital.
  - If a new X-ray tube is fitted or any major works are carried out on the scanner, please inform Mount Vernon Hospital Medical Physics Department and carry out QC tests.

TABLE 29 Radiographer phantom QC protocol

Scan parameter	Setting
Tube voltage	100 kVp
Tube current	350 mA
Rotation time	1.0 second or similar, depending on scanner
Pitch	1 : 1 or similar, depending on scanner
Field of view	25 cm or similar, depending on scanner
Z-direction coverage	At least 60 mm
Detector collimation	To be specified for each scanner model
Slice thickness	3.0 mm
Reconstruction interval	2.0 mm or similar, depending on scanner
Reconstruction algorithm	To be specified for each scanner model. Iterative reconstruction (if available) to be switched off



---

FIGURE 23 Measurement positions in QC phantom.

## Appendix 4 Number of scans that deviated from the scan protocol

TABLE 30 Number of scans that deviated from the scan protocol

Type of deviation (protocol/GCP/SOP)	Deviation relating to [protocol version (+ section)/GCP section/SOP number]	Deviation category	Deviation details
Protocol	Appendix 3	DCE-CT no scan	Nodule was not seen on the DCE-CT scan, so no analysis could be performed
Protocol	Appendix 3	DCE-CT reconstruction	Reconstruction interval 2.5 mm in the pre-contrast scan, rather than 2.0 mm
Protocol	Appendix 3	DCE-CT mA	From NCRI PET Core Lab QA, DCE-CT was performed at 500 mA, but should have been 350 mA because patient weight was not > 90 kg. Discussed at TMG: considered a minor deviation with minor impact
Protocol	Appendix 3	DCE-CT tube current	Tube current was 350 mA, but, for patient weight, it should have been 200 mA
Protocol	Appendix 3	DCE-CT slice thickness	The DCE-CT scan contained only 17 slices; other scans had contained 31–35 slices. JM asked CISC for an explanation of the reduced number of slices. CISC confirmed that the SPN was completely covered in the 17 slices
Protocol	Appendix 3	DCE-CT mA	The wrong current for patient weight was used
Protocol	Appendix 3	DCE-CT no scan	Withdrawn: patient did not have DCE-CT scan as was too breathless lying down in scanner following PET. Withdrew consent following PET
Protocol	Appendix 3	DCE-CT scan location	Operator error – the 60-second post-contrast scan was in the wrong location, so no SPN image was recorded
Protocol	Appendix 3	DCE-CT no scan	Patient has completely withdrawn consent
Protocol	Appendix 3	DCE-CT no scan	DCE-CT not carried out as nodule has resolved
Protocol	Appendix 3	DCE-CT time	From NCRI PET Core Lab QA: there was a delay on the DCE-CT scan – operator had paused CT acquisition for ≈ 23 seconds following the first post-contrast acquisition, so there is a ≈ 23-second delay on the 120-second, 180-second and 240-second scans. Discussed at a TMG meeting. This is a major protocol deviation that compromises the data. It is to be flagged to statistician at the time of data analysis
Protocol	Appendix 3	DCE-CT no scan	Patient has not undergone DCE-CT, but is happy to be followed up and to fill in both IPCARD-SPN questionnaires
Protocol	Appendix 3	DCE-CT no scan	Patient had panic attack and refused both DCE-CT and PET/CT, but did not withdraw consent

continued

TABLE 30 Number of scans that deviated from the scan protocol (continued)

Type of deviation (protocol/GCP/SOP)	Deviation relating to [protocol version (+ section)/GCP section/SOP number]	Deviation category	Deviation details
Protocol	Appendix 3	DCE-CT slice thickness	Slice thickness in DICOM header is 0.29 mm; protocol says should be 3 mm. NCRI PET Core Lab think this is an error as the slice separation is 2 mm and the images seem OK  Field of view not indicated in DICOM header, but was measured by NCRI PET Core Lab to be correct 150 mm
Protocol	Appendix 3	DCE-CT no scan	Patient agitated following PET; refused to stay for DCE-CT. Could not rebook DCE-CT in time. Did not withdraw consent
Protocol	Appendix 3	DCE-CT no scan	Patient had panic attack and refused contrast injection. Did not withdraw consent
Protocol	Appendix 3	DCE-CT mA	A variable current was used for DCE-CT. JM e-mailed radiologist Andy Scarsbrook, who enquired from radiography, who replied to say it was a one-off mistake
Protocol	Appendix 3	DCE-CT no scan	Patient underwent repeat CT. Radiologist decided SPN = benign; therefore, no PET follow-up was required. Patient did not withdraw consent from IPCARD-SPN questionnaire substudy
Protocol	Appendix 3	DCE-CT extra pre scan	Scan was sent to NCRI PET Core Lab with two sets of pre-contrast images
Protocol	Appendix 3	DCE-CT no scan	DCE-CT not carried out because of renal insufficiency
Protocol	Appendix 3	DCE-CT no scan	Refused both scans on day of appointment
Protocol	Appendix 3	DCE-CT no scan	PET was performed, but clinician decided against DCE as some infection within tissue did not look good as a result of poor venous access and thrombophlebitis
Protocol	Appendix 3	DCE-CT no scan	Failed cannular as contrast was being initiated; patient would not agree to re-cannulation. Apparently she had very poor access and was sore after the PET glucose was administered; DCE-CT not carried out
Protocol	Appendix 3	DCE-CT no scan	Patient has completely withdrawn consent
Protocol	Appendix 3	DCE-CT injection rate	Patient did not have DCE-CT scan within 14 days
Protocol	Appendix 3	DCE-CT field of view	The field of view for DCE-CT was 303 mm; protocol states 150 mm
Protocol	Appendix 3	DCE-CT injection rate	Contrast was injected at 3 ml/second, instead of 2 ml/second
Protocol	Appendix 3	DCE-CT injection rate	Contrast was injected at 3 ml/second, instead of 2 ml/second
Protocol	Appendix 3	DCE-CT injection rate	Contrast was injected at 3 ml/second, instead of 2 ml/second
Protocol	Appendix 3	DCE-CT injection rate	Contrast was injected at 3 ml/second, instead of 2 ml/second

TABLE 30 Number of scans that deviated from the scan protocol (continued)

Type of deviation (protocol/GCP/SOP)	Deviation relating to [protocol version (+ section)/GCP section/SOP number]	Deviation category	Deviation details
Protocol	Appendix 3	DCE-CT no scan	Patient's weight too great for scanner
Protocol	Appendix 3	DCE-CT nodule analysis	Owing to the large cavity of the nodule, it could not be analysed by setting a circular or oval region that covers two-thirds of the nodule. Chief investigator suggested using a horseshoe-shaped region, which could be set within the solid part. Elaine Smith (research nurse) thought this was feasible and could cover half of the solid part of the nodule. Therefore, the patient was included in the trial
Protocol	Appendix 3	DCE-CT no scan	DCE-CT not carried out as nodule reduced in size
Protocol	Appendix 3	DCE-CT reconstruction	From NCRI PET Core Lab QA: the wrong reconstruction diameter was used for the pre-contrast scan of NC753
Protocol	Appendix 3	DCE-CT reconstruction	From NCRI PET Core Lab QA: the wrong reconstruction diameter was used for all images for NC755
Protocol	Appendix 3	DCE-CT reconstruction	From NCRI PET Core Lab QA: the wrong reconstruction diameter was used for all images for NC756
Protocol	Appendix 3	DCE-CT pre scan	The DCE-CT scan had two pre-contrast scans of different areas of the lung
Protocol	Appendix 3	DCE-CT no scan	Patient did not undergo DCE-CT because of high creatinine levels
Protocol	Appendix 2	DCE-CT reconstruction	From NCRI PET Core Lab QA: the reconstruction diameter on the pre-contrast DCE-CT scan was 426 mm, instead of 150 mm
Protocol	Appendix 3	DCE-CT no scan	Withdrawn: patient underwent pre-contrast CT - nodule had reduced in size so no contrast and scan given. They did not withdraw consent. No follow-up for this patient
Protocol	Appendix 3	DCE-CT no scan	Withdrawn: patient had panic attack and refused imaging test; they did not withdraw consent. No follow-up for this patient
Protocol	Appendix 3	DCE-CT reconstruction	DCE-CT scan was reconstructed incorrectly: BR40s kernel instead of B30s kernel; e-mail confirming that BR40s was expected in site file
Protocol	Appendix 3	DCE-CT reconstruction	DCE-CT scan was reconstructed incorrectly: BR40s kernel instead of B30s kernel; e-mail confirming that BR40s was expected in site file
Protocol	Appendix 3	DCE-CT no scan	Withdrawn consent, due to 'cannulation'
Protocol	Appendix 3	DCE-CT reconstruction	From NCRI PET Core Lab QA: the post-contrast 2-, 3- and 4-minute scans had a reconstruction diameter of 153 mm; should be 150 mm. Discussed at TMG meeting: minor deviation with minimal impact on data

continued

TABLE 30 Number of scans that deviated from the scan protocol (continued)

Type of deviation (protocol/GCP/SOP)	Deviation relating to [protocol version (+ section)/GCP section/SOP number]	Deviation category	Deviation details
Protocol	Appendix 3	DCE-CT reconstruction	From NCRI PET Core Lab QA: DCE-CT scan had reconstruction interval of 3 mm for the pre-contrast scan (should be 2 mm or similar, depending on scanner). Discussed at TMG: minor deviation with minimal impact on data
Protocol	Appendix 3	DCE-CT reconstruction	From NCRI PET Core Lab QA: DCE-CT scan had reconstruction interval of 3 mm for the pre-contrast scan (should be 2 mm or similar, depending on scanner). Discussed at TMG: minor deviation with minimal impact on data
Protocol	Appendix 3	DCE-CT reconstruction	From NCRI PET Core Lab QA: DCE-CT scan had reconstruction interval of 3 mm for the pre-contrast scan (should be 2 mm or similar, depending on scanner). Discussed at TMG: minor deviation with minimal impact on data
Protocol	Appendix 3	DCE-CT slice thickness	From NCRI-PET Core Lab QA: 2-mm slice thickness on all slices; protocol states 3mm. Discussed at TMG: minor deviation with minimal impact on data
Protocol	Appendix 3	DCE-CT mA	Current used was 200 mA, but should have been 350 mA for weight of 64.3 kg
Protocol	Appendix 3	DCE-CT reconstruction	Scans were reconstructed using window width of 350 HU instead of 400 HU
Protocol	Appendix 3	DCE-CT window	Incorrect window width: 350 HU not 400 HU
Protocol	Appendix 3	DCE-CT reconstruction	From NCRI PET Core Lab QA: reconstruction diameter was 302 mm instead of 150 mm. Wrong diameter used in error as other scans are fine. It is unlikely that it can be reconstructed to the correct diameter
Protocol	Appendix 3	DCE-CT window	From NCRI PET Core Lab QA: incorrect window width – 350 HU not 400 HU
Protocol	Appendix 3	DCE-CT no scan	Patient did not undergo DCE-CT as nodule had resolved and reduced in size at DCE-CT appointment
Protocol	Appendix 3	DCE-CT tube current	Tube current was 350 mA, but for patient weight (99 kg) it should have been 500 mA. Site informed and asked to remind radiographers to scan to protocol
Protocol	Appendix 3	DCE-CT no scan	Withdrawn consent
Protocol	Appendix 3	DCE-CT QA	Monitoring: DCE-CT QA scan data have not been sent for quality checks. Scan data have been sent electronically to Mount Vernon Hospital Medical Physics Department
Protocol	Appendix 3	DCE-CT no scan	Nodule too small to analyse on DCE-CT on pre-locating scan. Did not withdraw consent
Protocol	Appendix 3	DCE-CT no scan	Nodule too small to analyse on DCE-CT on pre-locating scan. Did not withdraw consent
Protocol	Appendix 3	DCE-CT no scan	Nodule too small to analyse on DCE-CT on pre-locating scan. Did not withdraw consent

TABLE 30 Number of scans that deviated from the scan protocol (*continued*)

Type of deviation (protocol/GCP/SOP)	Deviation relating to [protocol version (+ section)/GCP section/SOP number]	Deviation category	Deviation details
Protocol	Appendix 3	DCE-CT no scan	Nodule too small to analyse on DCE-CT on pre-locating scan. Did not withdraw consent
Protocol	Inclusion/exclusion criteria ( <i>Chapter 2, Participants</i> )	DCE-CT no scan	Patient appeared eligible on CT scan, but nodule could not be located for DCE-CT
Protocol	Appendix 3	DCE-CT no scan	No DCE-CT carried out as consent not sent to CISC on time
Protocol	Appendix 3	DCE-CT no scan	No DCE-CT carried out as consent not sent to CISC on time

CISC, Clinical Imaging Sciences Centre; DICOM, Digital Imaging and Communications in Medicine; GCP, good clinical practice; NCRI, National Cancer Research Institute; SOP, standard operating procedure.



# Appendix 5 Diagnostic accuracy: additional tables

## Search strategy

Database: EMBASE.

Date range searched: 1974 to 28 February 2019.

Database: Ovid MEDLINE(R).

Date range searched: 1946 to 28 February 2019.

Date searched: 28 February 2019.

1. Solitary Pulmonary Nodule.mp. [mp=ti, ab, hw, tn, ot, dm, mf, dv, kw, fx, dq, nm, kf, px, rx, an, ui, sy] (5712)
2. SPN.tw. (4399)
3. (pulmonary or lung or chest or pleura).tw. (2,368,476)
4. 2 and 3 (1041)
5. ((solitary or coin or single or discrete or indeterminate) adj3 (lung or pulmonary or "chest wall" or pleura) adj3 (lesion\* or lump\* or nodule\* or lobe or lobes)).tw. (5990)
6. 1 or 4 or 5 (8912)
7. "Computed Tomography"/ (677,626)
8. Tomography, X-Ray Computed/ (390,442)
9. Tomography, Spiral Computed/ (18,211)
10. Contrast Media/ (133,577)
11. (Dynamic adj3 enhance\*).tw. (20,256)
12. (contrast adj3 enhance\*).tw. (118,030)
13. "DCE-CT".tw. (423)
14. (DCE and CT).tw. (1154)
15. "perfusion".tw. (351,014)
16. "computed tomography".tw. (485,681)
17. 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 (1,750,697)
18. 6 and 17 (4539)
19. exp accuracy/ (160,921)
20. (diagnos\* or sensitiv\* or specific\* or accura\*).tw. (13,447,155)
21. differenti\*.tw. (2,568,154)
22. 19 or 20 or 21 (14,882,434)
23. (letter or editorial or comment or historical article).pt. (3,666,677)
24. 22 not 23 (14,742,125)
25. 18 and 24 (3148)
26. diagnosis, computer-assisted/ or image interpretation, computer-assisted/ (100,696)
27. 6 and 24 and 26 (221)
28. 25 or 27 (3211)
29. limit 28 to humans (3028).

TABLE 31 Summary of the study design and baseline characteristics of those included in the meta-analysis

Study	Year	Country	Design	Centres (n)	Population size (n)	Mean age (years)	Mean nodule size (mm) (range)	Reference standard	Nodule diagnosis (M/B)
Swensen <i>et al.</i> <sup>90</sup>	1992	USA	Retrospective	1	30	60	16.4 (6–30)	Histology or follow-up	23 M/7B
Swensen <i>et al.</i> <sup>99</sup>	1995	USA	Retrospective	1	163	63	17.8 (6–40)	Histology or follow-up	111 M/52B
Yamashita <i>et al.</i> <sup>100</sup>	1995	Japan	Retrospective	1	32	52	16.7 (2–30)	Histology	18 M/14 B
Swensen <i>et al.</i> <sup>101</sup>	1996	USA	Retrospective	1	107	63	X (7–30)	Histology or follow-up	52 M/55 B
Potente <i>et al.</i> <sup>102</sup>	1997	Italy	Retrospective	1	25	64	18.2 (5–30)	Histology	17 M/8 B
Zhang and Kono <sup>103</sup>	1997	Japan	Retrospective	1	65	64	19.1 (5–30)	Histology or follow-up	42 M/23 B
Swensen <i>et al.</i> <sup>15</sup>	2000	USA	Prospective	7	356	64	15.3 (5–40)	Histology or follow-up	171 M/184 B
Kim <i>et al.</i> <sup>104</sup>	2004	Republic of Korea	Prospective	1	50	50	21 (7–38)	Histology or follow-up	19 M/31 B
Orlacchio <i>et al.</i> <sup>32</sup>	2007	Italy	Prospective	1	56	63	X (X–30)	Histology or follow-up	26 M/30 B
Lee <i>et al.</i> <sup>89</sup>	2007	Republic of Korea	Prospective	1	486	56	19.6 (5.5–30)	Histology or follow-up	237 M/249 B
Ohno <i>et al.</i> <sup>34</sup>	2008	Japan	Prospective	1	175	72	15.7 (8–29)	Histology or follow-up	152 M/50 B
Choi <i>et al.</i> <sup>105</sup>	2008	Republic of Korea	Retrospective	1	40	56	20.6 (12–30)	Histology	13 M/27 B
Bayraktaroglu <i>et al.</i> <sup>106</sup>	2008	Turkey	Retrospective	1	22	50	20 (10–35.5)	Histology or follow-up	9 M/13 B
Bai <i>et al.</i> <sup>107</sup>	2009	China	Prospective	1	68	53	23 (8–30)	Histology	36 M/32 B
Jiang <i>et al.</i> <sup>108</sup>	2009	China	Retrospective	1	51	50	26.5 (10–40)	Histology or follow-up	28 M/23 B
Dabrowska <i>et al.</i> <sup>35</sup>	2010	Poland	Retrospective	1	40	61	20.3 (10–40)	Histology or follow-up	23 M/17 B
Li <i>et al.</i> <sup>109</sup>	2010	China	Prospective	1	77	56	X (X–30)	Histology	46 M/22 B
Ohno <i>et al.</i> <sup>36</sup>	2011	Japan	Prospective	1	50	74	15.8 (4–29)	Histology or follow-up	43 M/33 B
Ohno <i>et al.</i> <sup>37</sup>	2013	Japan	Prospective	1	52	72	15.9 (4–29)	Histology or follow-up	57 M/39 B
Shu <i>et al.</i> <sup>110</sup>	2013	China	Prospective	1	144	53	23 (8–30)	Histology	76 M/68 B
Ribeiro <i>et al.</i> <sup>111</sup>	2013	Brazil	Retrospective	1	23	60	15 (5–30)	Histology or follow-up	5 M/18 B
Ye <i>et al.</i> <sup>112</sup>	2014	China	Prospective	1	87	59	17.2 (5–30)	Histology or follow-up	52 M/35 B
Ohno <i>et al.</i> <sup>33</sup>	2015	Japan	Prospective	1	198	75	18.4 (8–29)	Histology or follow-up	133 M/85 B

B, benign; M, malignant; X, not reported in paper.

TABLE 32 Summary of the CT acquisition protocols and measurements of the studies included in the meta-analysis

Study	CT technique <sup>a</sup>	Contrast	Contrast volume and rate	Slice thickness (mm)	Scan timing (seconds) <sup>b</sup>	kV	mA	Enhancement threshold(s) <sup>c</sup>	Threshold set prospectively
Swensen <i>et al.</i> <sup>90</sup>	1	Omnipaque 300 (GE Healthcare)	100 ml at 2 ml/second	1.5/2	0, 60, 120, 180, 240, 300	<sup>a</sup>	<sup>a</sup>	20	No
Swensen <i>et al.</i> <sup>99</sup>	1	Omnipaque 300	100 ml at 2 ml/second	1.5/2/3	0, 60, 120, 180, 240, 300	120/130	<sup>a</sup>	20	No
Yamashita <i>et al.</i> <sup>100</sup>	1	Omnipaque 300	100–150 ml at 2 ml/second	2	0, 30, 120, 300	<sup>a</sup>	<sup>a</sup>	20	No
Swensen <i>et al.</i> <sup>101</sup>	1	ISOVUE® (Bracco S.p.A., Milan, Italy)	420 mgI/kg at 2 ml/second	1/2	0, 60, 120, 180, 240, 300	120/130	280/200	20	Yes
Potente <i>et al.</i> <sup>102</sup>	1	Omnipaque 300	450 mgI/kg at 2ml/second	1	0, 60, 120, 180, 240, 300	120/140	<sup>a</sup>	20	No
Zhang and Kono <sup>103</sup>	1	Iopamiron® (Bracco S.p.A.)	100 ml at 4 ml/second	5	20–24 images over 105–165 seconds	<sup>a</sup>	<sup>a</sup>	20	No
Swensen <i>et al.</i> <sup>15</sup>	1	<sup>a</sup>	420 mgI/kg at 2 ml/second	2	0, 60, 120, 180, 240, 300	120	280	15	Yes
Kim <i>et al.</i> <sup>104</sup>	1	Omnipaque 300	80 ml at 2.5 ml/second	2	0, 60, 120, 180, 240, 300	120	160	20	Yes
Orlacchio <i>et al.</i> <sup>32</sup>	1	<sup>a</sup>	420mgI/kg at 3 ml/second	1.25	0, 60–80	120/140	Auto mA	15	Yes
Lee <i>et al.</i> <sup>89</sup>	1	Iomeprol (Bracco, Milan, Italy)	120 ml at 3 ml/second	2.5	0, 30, 60, 90, 120, 240, 300, 540, 720, 900	120	90	25/25 wash-in, 5–36 wash-out	No
Ohno <i>et al.</i> <sup>34</sup>	1	Iopamiron (Bracco, Milan, Italy)	100 ml at 4 ml/second	2	0, 30, 60, 90, 120, 300	120	60	≥ 20 HU wash-in, < 30 HU wash-out	No
Choi <i>et al.</i> <sup>105</sup>	1	Ultravist® (Bayer AG, Leverkusen, Germany)	120 ml at 3 ml/second	2	0, 20, 40, 60, 80, 120, 140, 160, 180, 240, 300	120	170	15/≥ 15 HU wash-in, 5–25 HU wash-out	Yes
Bayraktaroglu <i>et al.</i> <sup>106</sup>	1		100 ml at 3 ml/second	2	0, 60, 120, 180, 240, 300	140	120–140	15/20	Yes

continued

TABLE 32 Summary of the CT acquisition protocols and measurements of the studies included in the meta-analysis (continued)

Study	CT technique <sup>a</sup>	Contrast	Contrast volume and rate	Slice thickness (mm)	Scan timing (seconds) <sup>b</sup>	kV	mA	Enhancement threshold(s) <sup>c</sup>	Threshold set prospectively
Bai <i>et al.</i> <sup>107</sup>	1	Ultravist	100 ml at 4 ml/second	1	Five sets of images acquired at 0, 15, 75, 135, 193 and 251 seconds. The first and second series were scanned 10 times (each scanning duration 1 second and scanning interval 2 seconds). The third, fourth and fifth series were scanned four times (each scanning duration 1 second and scanning interval 6 seconds). The delayed scanning interval between each series was 30 seconds	120	300	20	Yes
Jiang <i>et al.</i> <sup>108</sup>	1	Ultravist	1.5 ml/kg at 3.2 ml/second	2	0, 15, 45, 75, 135, 195, 255	120	125	15/20/25	Yes
Dabrowska <i>et al.</i> <sup>35</sup>	1	Iomeprol	420 mgI/kg at 2 ml/second	3	0, 30, 240	120/140	250	15/20/30	No
Li <i>et al.</i> <sup>109</sup>	2	Ultravist	50 ml at 6–7 ml/second	3	55-second scan time with 0.4-second rotation time	120	100	23.3 HU/12.2 ml per 100 g	No
Ohno <i>et al.</i> <sup>36</sup>	2	Iopamiron	0.2 ml/kg at 5 ml/second	2	0, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 28, 40, 50, 60, 90, 120	80	120	40 ml/100 ml/minute	No
Ohno <i>et al.</i> <sup>37</sup>	2	Iopamiron	0.2 ml/kg at 5 ml/second	2	0, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 28, 40, 50, 60, 90, 120	80	120	40 ml/100 ml/minute	No
Shu <i>et al.</i> <sup>110</sup>	2	Ultravist	50 ml at 5 ml/second	5	40 seconds, starting at 0 seconds with 2-second intervals	120	60	6 ml/100 g	No
Ribeiro <i>et al.</i> <sup>111</sup>	1		420 mgI/kg at 2 ml/second	3	0, 180, 240, 300	120	190	15	Yes
Ye <i>et al.</i> <sup>112</sup>	1	Iopamidol	420 mgI/kg at 4 ml/second	2	0, 20, 30, 45, 60, 75, 90, 120, 180, 300, 540, 720, 900, 1200	120	240	0.018% wash-out per second	No
Ohno <i>et al.</i> <sup>33</sup>	2	Iopamiron	0.5 ml/kg at 5 ml/second	2	0, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 28, 40, 50, 60, 90, 120	80	120	29 ml/100 ml/minute	No

a Scan type as follows: 1 – mono-energetic DCE-CT; 2 – CT perfusion; 3 – dual-energy DCE-CT.

b Dual-energy scans allow for generation of a simulated non-contrast scan and for enhancement quantification based on this. When this has occurred, it has been recorded as (0).

c The HU difference between baseline and peak enhancement, unless otherwise stated.

TABLE 33 The QUADAS-2 components for each of the individual studies

Study	Risk of bias				Applicability concerns		
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Swensen <i>et al.</i> <sup>90</sup>	?	⊖	?	⊖	?	⊕	⊕
Swensen <i>et al.</i> <sup>99</sup>	?	⊖	⊕	?	?	⊕	⊕
Yamashita <i>et al.</i> <sup>100</sup>	?	⊖	?	?	⊕	⊕	⊕
Swensen <i>et al.</i> <sup>101</sup>	?	⊕	⊕	?	⊕	⊕	⊕
Potente <i>et al.</i> <sup>102</sup>	⊖	?	?	⊖	⊕	⊕	⊕
Zhang and Kono <sup>103</sup>	?	⊖	?	?	⊕	⊕	⊕
Swensen <i>et al.</i> <sup>15</sup>	?	⊕	?	?	⊕	⊕	⊕
Kim <i>et al.</i> <sup>104</sup>	⊖	⊕	?	⊖	⊖	?	⊕
Orlacchio <i>et al.</i> <sup>32</sup>	?	?	⊖	?	?	⊕	?
Lee <i>et al.</i> <sup>89</sup>	?	⊖	?	⊖	?	⊕	⊕
Ohno <i>et al.</i> <sup>34</sup>	⊕	⊖	?	?	⊖	⊕	⊕
Choi <i>et al.</i> <sup>105</sup>	⊖	⊕	?	?	⊕	⊕	⊕
Bayraktaroglu <i>et al.</i> <sup>106</sup>	?	⊕	?	⊖	⊕	⊕	⊕
Bai <i>et al.</i> <sup>107</sup>	?	?	?	?	⊕	⊕	⊕
Jiang <i>et al.</i> <sup>108</sup>	?	?	?	?	⊕	?	⊕
Dabrowska <i>et al.</i> <sup>35</sup>	⊖	⊖	?	⊖	⊖	⊕	⊕
Li <i>et al.</i> <sup>109</sup>	?	⊖	?	?	⊕	⊕	⊕
Ohno <i>et al.</i> <sup>36</sup>	⊕	⊖	?	⊕	⊖	⊕	⊕
Ohno <i>et al.</i> <sup>37</sup>	⊕	⊖	?	⊕	⊖	⊕	⊕
Shu <i>et al.</i> <sup>110</sup>	?	⊖	?	⊕	?	?	⊕
Ribeiro <i>et al.</i> <sup>111</sup>	⊕	⊕	?	?	⊕	⊕	⊕
Ye <i>et al.</i> <sup>112</sup>	?	⊖	?	?	⊕	⊕	⊕
Ohno <i>et al.</i> <sup>33</sup>	⊕	⊖	?	⊕	⊕	⊕	⊕

⊕, Low risk of bias; ⊖, high risk of bias; ?, unclear risk of bias.

TABLE 34 Diagnostic performance of DCE-CT for the evaluation of pulmonary nodules

Threshold	Studies (n)	Patients (n)	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	PLR (95% CI)	NLR (95% CI)	DOR (95% CI)
<b>All</b>							
	23	2397	94.8 (91.5 to 96.9)	75.5 (69.4 to 80.6)	3.86 (2.99 to 4.74)	0.07 (0.03 to 0.10)	56.6 (24.2 to 88.9)
<b>Enhancement thresholds</b>							
15 HU	7	588	97.2 (93.9 to 98.8)	64.3 (42.4 to 81.5)	2.72 (1.18 to 4.27)	0.04 (0.01 to 0.07)	63.5 (5.2 to 121.8)
20 HU	11	653	98.3 (95.1 to 99.4)	71.0 (63.1 to 77.8)	3.39 (2.50 to 4.28)	0.02 (0.00 to 0.05)	142.5 (−36.4 to 321.3)

TABLE 35 Subgroup analyses of the diagnostic performance of DCE-CT for the evaluation of indeterminate pulmonary lesions

Characteristic	Studies (n)	Patients (n)	Lesions (n)	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	p-value
<b>CT technique</b>						
MECT	18	1876	1903	95.7 (91.9 to 97.8)	74.6 (67.1 to 80.9)	0.42
CTP	5	521	611	91.5 (88.2 to 94.0)	78.7 (71.7 to 84.3)	
<b>Sample size</b>						
< 100	16	768	838	94.5 (89.3 to 97.3)	78.6 (69.7 to 85.5)	0.55
≥ 100	7	1629	1676	95.1 (91.0 to 97.4)	71.4 (64.3 to 77.6)	
<b>Mean lesion size<sup>a</sup></b>						
< 20 mm	13	1742	1859	95.4 (89.5 to 98.1)	72.7 (66.4 to 78.2)	0.99
≥ 20 mm	7	415	415	93.1 (87.7 to 96.3)	72.6 (59.8 to 82.6)	
<b>Maximum lesion size<sup>b</sup></b>						
≤ 30 mm	17	1715	1832	93.1 (88.9 to 95.9)	78.0 (70.8 to 83.8)	0.07
> 30 mm	6	682	682	98.0 (93.3 to 99.4)	67.9 (57.6 to 76.7)	
<b>Threshold prospectively set</b>						
Yes	8	723	723	95.7 (92.5 to 97.7)	77.2 (60.9 to 88.0)	0.84
No/unclear	15	1791	1791	94.9 (89.7 to 97.6)	75.3 (70.2 to 79.9)	
<b>Patient selection bias</b>						
Low	6	538	655	93.0 (89.2 to 95.5)	67.0 (57.3 to 75.4)	0.14
Yes/unclear	17	1859	1859	95.4 (91.2 to 97.7)	78.7 (71.6 to 84.4)	
<b>Index test bias</b>						
Low	6	598	598	96.0 (91.5 to 98.2)	70.6 (59.7 to 79.6)	0.62
Yes/unclear	17	1859	1916	94.7 (90.3 to 97.2)	77.0 (69.9 to 82.9)	
<b>Reference standard bias</b>						
Low	2	270	270	99.4 (93.5 to 99.9)	74.8 (65.6 to 82.2)	0.044
Yes/unclear	21	2127	2244	93.6 (90.0 to 95.9)	75.5 (68.7 to 81.3)	
<b>Flow and timing bias</b>						
Low	4	444	534	91.3 (87.6 to 93.9)	77.0 (70.5 to 82.4)	0.63
Yes/unclear	19	1953	1980	95.7 (91.9 to 97.8)	75.4 (67.5 to 81.8)	
<b>Publication date</b>						
Pre 2008	10	1370	1370	97.2 (93.2 to 98.8)	77.8 (67.4 to 85.5)	0.07
2008 onwards	13	1027	1144	92.0 (86.5 to 95.4)	73.8 (65.8 to 80.4)	
CTP, computerised tomography perfusion; MECT, mono-energetic dynamic contrast enhanced computerised tomography. a Mean lesion size not reported in three studies, and reported as volumes rather than diameter in three studies. b Maximum lesion size reported as volumes in three studies.						

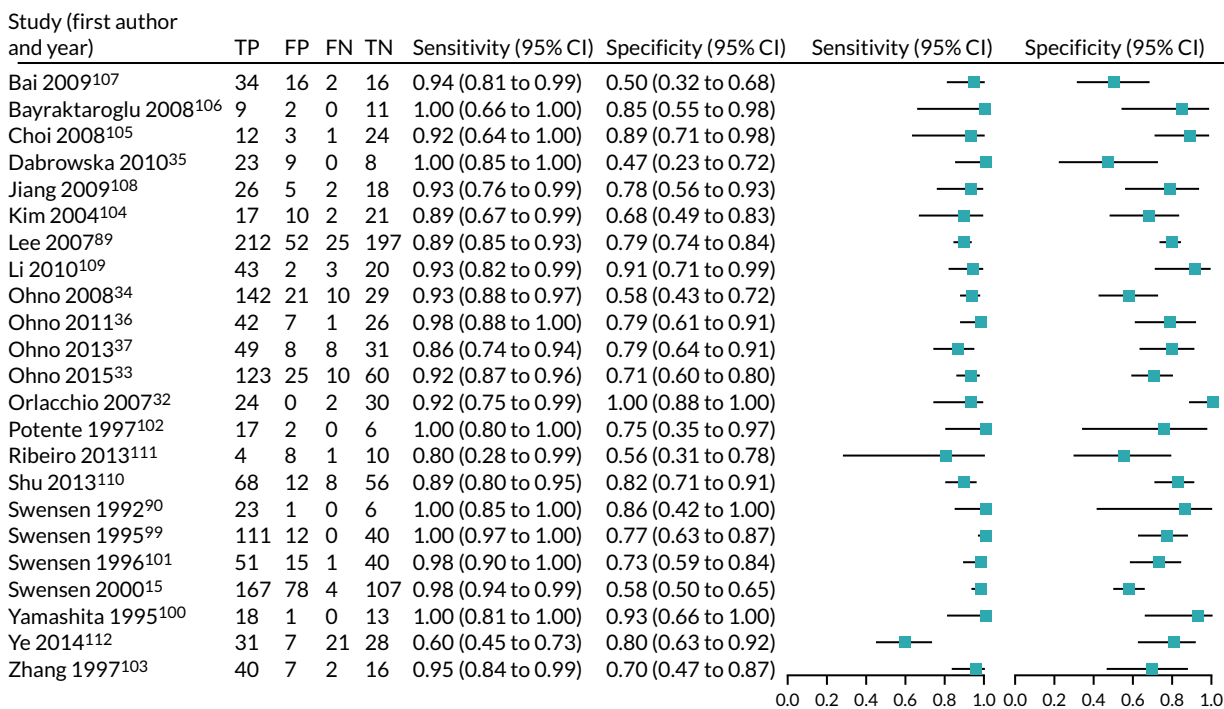


FIGURE 24 Forest plot of the included studies. Studies listed by first author and year of publication. FN, false negative; FP, false positive; TN, true negative; TP, true positive.

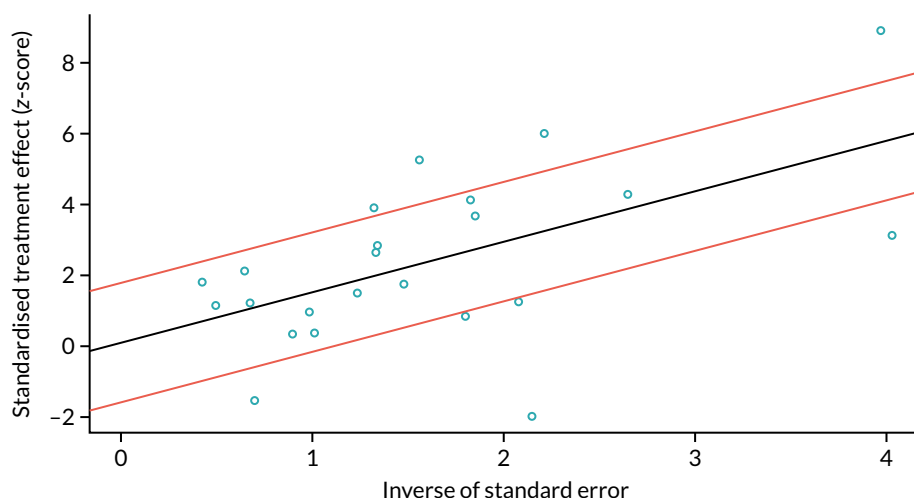


FIGURE 25 Galbraith plot examining interstudy heterogeneity for publication bias by incorporating the effect size of each study, compared with the pooled analysis. The y-axis represents the test statistics (effect/standard error of the estimate) of each study, which are expected to fall within 2 units of the pooled effects for 95% of the studies. The x-axis plots 1/standard error of the pooled study estimate.



# Appendix 6 MEDLINE search strategy

## Search strategy

Database: Ovid MEDLINE(R).

Date range searched: 1946 to July week 1 2013.

Date search was conducted: 1 March 2019.

1. Solitary Pulmonary Nodule/ (2896)
2. SPN.tw. (1048)
3. (pulmonary or lung or chest or pleura).tw. (755,742)
4. 2 and 3 (211)
5. ((solitary or coin or single or discrete or indeterminate) adj3 (lung or pulmonary or "chest wall" or pleura) adj3 (lesion\* or lump\* or nodule\* or lobe or lobes)).tw. (1846)
6. 1 or 4 or 5 (3820)
7. "Positron-Emission Tomography and Computed Tomography"/ (2514)
8. Positron-Emission Tomography/ (27,996)
9. Tomography, X-Ray Computed/ (277,285)
10. Tomography, Spiral Computed/ (6448)
11. Fluorodeoxyglucose F18/ (17,953)
12. Fluorine Radioisotopes/ (4789)
13. fluorodeoxyglucose\*.tw. (7890)
14. Contrast Media/ (65,791)
15. (FDG and PET and CT).tw. (6401)
16. "FDG-PET-CT".tw. (3282)
17. "DCE-CT".tw. (81)
18. (DCE and CT).tw. (169)
19. "positron emission tomography".tw. (31,770)
20. "computed tomography".tw. (126,736)
21. or/7-20 (423,484)
22. 6 and 21 (1987)
23. exp economics/ (486,868)
24. exp economics hospital/ (19,129)
25. exp economics pharmaceutical/ (2573)
26. exp economics nursing/ (3874)
27. exp economics medical/ (13,457)
28. exp "Costs and Cost Analysis"/ (179,830)
29. Cost Benefit Analysis/ (60,247)
30. exp models economic/ (10,064)
31. exp fees/ and charges/ (8147)
32. exp budgets/ (11,907)
33. (economic\* or cost or costs or costly or costing or price or prices or pricing or pharmaco-economic\*).tw. (416,680)
34. (value adj1 money).tw. (22)
35. budget\$.tw. (16,959)
36. or/23-35 (770,608)
37. ((energy or oxygen) adj cost).tw. (2647)
38. (metabolic adj cost).tw. (747)
39. ((energy or oxygen) adj expenditure).tw. (16,124)

40. or/37-39 (18,835)
41. 36 not 40 (766,435)
42. (letter or editorial or comment or historical article).pt. (1,500,002)
43. 41 not 42 (707,321)
44. 22 and 43 (106)
45. mass screening/ or diagnostic techniques radioisotopes/ or diagnostic techniques respiratory sytem/ (81,834)
46. 6 and 43 and 45 (8)
47. 44 or 46 (107)
48. diagnosis, computer-assisted/ or image interpretation, computer-assisted/ (48,688)
49. 6 and 43 and 48 (6)
50. 47 or 49 (109).

## Appendix 7 Critical appraisal checklist

Item	Study identifier
1. Is there a clear statement of the decision problem?	
2. Is the comparator routinely used in clinical practice?	
3. Is the perspective of the model clearly stated?	
4. Is the study type appropriate?	
5. Is the modelling methodology appropriate?	
6. (a) Is the model structure described?	
(b) Does it reflect the disease process?	
7. Are assumptions about model structure listed and justified?	
8. Are the data inputs for the model described and justified?	
9. Is the effectiveness of the intervention (diagnostic accuracy) established based on a systematic review?	
10. (a) Are health benefits measured in QALYs?	
(b) Are health benefits measured using a standardised and validated generic instrument?	
11. (a) Are resource inputs described and justified?	
(b) Are resources valued appropriately?	
12. Have the costs and outcomes been discounted?	
13. Has an incremental analysis of costs and consequences of alternatives been performed?	
14. Has uncertainty been adequately assessed?	



## Appendix 8 List of excluded studies

NHS Quality Improvement Scotland. *Positron emission tomography (PET) imaging in cancer management; HTA Advice 2: Positron emission tomography (PET) imaging in cancer management; understanding HTBS Advice; use of PET imaging for cancer in Scotland*. Amendment to full report published July 2005. Glasgow: Health Technology Board for Scotland (HTBS)/NHS Quality Improvement Scotland (NHS QIS). HTA Report 2. 2002.

Baldwin DR, Eaton T, Kolbe J, Christmas T, Milne D, Mercer J, *et al*. Management of solitary pulmonary nodules: how do thoracic computed tomography and guided fine needle biopsy influence clinical decisions? *Thorax* 2002;**57**:817–22 (study design).

Cao JQ, Rodrigues GB, Louie AV, Zaric GS. Systematic review of the cost-effectiveness of positron-emission tomography in staging of non-small-cell lung cancer and management of solitary pulmonary nodules. [review]. *Clin Lung Cancer* 2012;**13**:161–70 (study design).

Carbone RG, Musi M, Peano L, Sblendorio L, Cantalupi D, Bossone E. Positron emission (PET) versus computed tomography (CT) in indeterminate pulmonary nodules (PN) diagnosis and mediastinal staging: cost-effectiveness analysis in Italy. *Chest* 2000;**118**(Suppl. 1):230S (abstract).

Dewan N, Reeb S, Gupta N, Lillington G, Scott W, O'Donohue WJJ, *et al*. Decision analysis to compare the cost of PET-FDG imaging with various treatment strategies for solitary pulmonary nodules. *Chest* 1994;**106**(Suppl. 1):88S (abstract).

Dwamena B, Fendrick A, Wahl R. Management of solitary pulmonary nodules using PDG-PET: decision and economic analyses. *J Nucl Med* 1996;**37**(Suppl. 1):111P (abstract).

Gambhir S, Shepherd J, Shah B, Schwimmer J, Czernin J, Phelps M. Cost effective analysis modeling for the role of FDG coincidence imaging (CI) in the management of patients with a solitary pulmonary nodule. *J Nucl Med* 1998;**39**(Suppl. 1):248P–9P (abstract).

Gerke O, Hermansson R, Hess S, Schifter S. Cost-effectiveness of PET and PET/computed tomography: a systematic review. *PET Clinics* 2015;**10**:105–24 (study design).

Goehler A, McMahon PM, Lumish HS, Wu CC, Munshi V, Gilmore M, *et al*. Cost-effectiveness of follow-up of pulmonary nodules incidentally detected on cardiac computed tomographic angiography in patients with suspected coronary artery disease. *Circulation* 2014;**130**:668–75 (population).

Gould MK, Lillington GA. Strategy and cost in investigating solitary pulmonary nodules. *Thorax* 1998;**53**:S32–7 (study design).

Lejeune C, Al ZK, Woronoff-Lemsi MC, Arveux P, Bernard A, Binquet C, *et al*. Use of a decision analysis model to assess the medicoeconomic implications of FDG PET imaging in diagnosing a solitary pulmonary nodule. *Eur J Health Econ* 2005;**6**:203–14 (only 1 year of follow-up).

Louie AV, Senan S, Patel P, Ferket BS, Lagerwaard FJ, Rodrigues GB, *et al*. When is a biopsy-proven diagnosis necessary before stereotactic ablative radiotherapy for lung cancer? A decision analysis. *Chest* 2014;**146**:1021–8 (intervention).

Murthy M, Medeiros T, Radhakrishnan J, Cardinal de Silva V, Irion K, Ledson M, *et al*. Incidental non-calcified pulmonary nodules: rationale for CT scanning and cost analysis. *Thorax* 2013;**68**:A98–9 (population).

Pyenson BS, Henschke CI, Yankelevitz DF, Yip R, Dec E. Offering lung cancer screening to high-risk medicare beneficiaries saves lives and is cost-effective: an actuarial analysis. *Am Health Drug Benefits* 2014;**7**:272–81 (comparator).

Romeo E, Gustavsen G, Buckingham J, Cole D, Narrow D, Sozzi G, *et al.* System economic impact of the miRNA signature classifier (MSC) test for management of patients with suspicious lung nodules. *Int J Radiat Oncol Biol Phys* 2014;**90**(Suppl. 1):S63 (abstract).

Rutter CE, Lester-Coll NH, Yu JB, Decker RH. Cost effectiveness of biopsy prior to stereotactic body radiation therapy (SBRT) for screening-detected FDG-avid lung nodules. *Int J Radiat Oncol Biol Phys* 2014;**90**(Suppl. 1):S137 (abstract).

Valk P, Hopkins D, Tesar R, Pounds T, Abella-Columna E, Haseman M, *et al.* Cost-effectiveness of PET imaging in management of solitary pulmonary nodules and non-small cell lung cancer. *J Nucl Med* 1996;**37**(Suppl.):111P (abstract).

Verboom P, van Tinteren H, Hoekstra OS, Smit EF, van den Bergh JHAM, Schreurs AJM, *et al.* Cost-effectiveness of FDG-PET in staging non-small cell lung cancer: the PLUS study. *Eur J Nucl Med Mol Imaging* 2003;**30**:1444–9 (study design).

Weber W, Buelow H, Roemer W, Praeuer H, Gambhir S, Schwaiger M. FDG-PET in solitary pulmonary nodules: a German cost-effectiveness analysis. *J Nucl Med* 1997;**38**(Suppl. 1):245P (abstract).

Weber W. Cost effectiveness of FDG-PET in the evaluation of solitary pulmonary nodules. *Eur J Cancer* 1997;**33**(Suppl. 9):S39 (abstract).

## Appendix 9 Systematic review of economic evaluations: tables

TABLE 36 Diagnostic pathways evaluated in included studies

Diagnostic technology	Study
<b>Comparator/baseline</b>	
No diagnostic test	Keith <i>et al.</i> <sup>21</sup> and Comber <i>et al.</i> <sup>26</sup>
Watch and wait <sup>a</sup>	Gambhir <i>et al.</i> , <sup>22</sup> Dietlein <i>et al.</i> , <sup>24</sup> Lejeune <i>et al.</i> <sup>28</sup> and Gould <i>et al.</i> <sup>20</sup>
CT <sup>b</sup>	Gugiatti <i>et al.</i> <sup>25</sup> and Tsushima and Endo <sup>27</sup>
PET/CT	Deppen <i>et al.</i> <sup>55</sup>
<b>Other comparators</b>	
Navigation bronchoscopy	Deppen <i>et al.</i> <sup>55</sup>
CT-guided biopsy	Gould <i>et al.</i> <sup>20</sup> and Deppen <i>et al.</i> <sup>55</sup>
Surgery <sup>c</sup>	Gambhir <i>et al.</i> , <sup>22</sup> Dietlein <i>et al.</i> <sup>24</sup> and Deppen <i>et al.</i> <sup>55</sup>
CT <sup>d</sup>	Gambhir <i>et al.</i> , <sup>22</sup> Keith <i>et al.</i> , <sup>21</sup> Comber <i>et al.</i> <sup>26</sup> and Gould <i>et al.</i> <sup>20</sup>
CT then biopsy	Gould <i>et al.</i> <sup>20</sup>
CT then PET <sup>e</sup>	Comber <i>et al.</i> <sup>26</sup> and Gould <i>et al.</i> <sup>20</sup>
CT and PET	Gould <i>et al.</i> <sup>20</sup>
PET <sup>f</sup>	Lejeune <i>et al.</i> <sup>28</sup>
PET then CT	Gould <i>et al.</i> <sup>20</sup>
<b>Intervention(s)</b>	
CT then PET <sup>e</sup>	Gambhir <i>et al.</i> , <sup>22</sup> Keith <i>et al.</i> , <sup>21</sup> Gugiatti <i>et al.</i> <sup>25</sup> and Tsushima and Endo <sup>27</sup>
CT and PET	Lejeune <i>et al.</i> <sup>28</sup>
PET <sup>f</sup>	Dietlein <i>et al.</i> <sup>24</sup>
CT then QECT <sup>g</sup>	Comber <i>et al.</i> <sup>26</sup>
CT then QECT then PET <sup>g</sup>	Comber <i>et al.</i> <sup>26</sup>
CT then CT-guided biopsy <sup>h</sup>	Tsushima and Endo <sup>27</sup>
CT then PET then CT-guided biopsy <sup>h</sup>	Tsushima and Endo <sup>27</sup>

a Serial chest X-ray (3, 6, 12, 24, 52, 104 weeks);<sup>22</sup> serial CT (3, 6, 12, 24, months);<sup>24</sup> serial CT (3, 6, 9, 12 months);<sup>28</sup> serial chest X-ray (1, 2, 4, 6, 9, 12, 15, 18, 21, 24 months).<sup>22</sup>

b Positive CT has CT-guided biopsy (20%) or thoracotomy (80%) – otherwise, discharge without follow-up;<sup>25</sup> positive CT followed by surgery, negative CT has single follow-up CT (positive follow-up CT has surgery, otherwise discharge).<sup>27</sup>

c Thoracotomy without histological verification.

d Positive/indeterminate CT has thoracotomy (80%) or biopsy (20%) (positive biopsies have thoracotomy – otherwise watch and wait); otherwise, watch and wait.<sup>21,22,26</sup>

e Positive/indeterminate CT has PET; positive PET has thoracotomy (80%) or biopsy (20%) (positive biopsies have thoracotomy, otherwise watch and wait); negative PET has biopsy (positive biopsies have thoracotomy, otherwise watch and wait); negative CT leads to watch and wait.<sup>22,26,27</sup> In Gugiatti *et al.*<sup>25</sup> only, patients with positive CT have PET (those with negative or indeterminate CT are discharged without follow-up); patients with positive CT but negative PET are also discharged without follow-up. In Lejeune *et al.*,<sup>28</sup> all patients have both CT and PET (positive CT/positive PET have surgery; positive CT/negative PET and negative CT/positive PET have biopsy; negative CT/negative PET lead to watch and wait).

f Positive PET has surgery (85%) or biopsy (15%) (positive biopsies have thoracotomy, otherwise watch and wait); negative PET leads to watch and wait.<sup>28</sup> Positive PET has surgery without histological verification; negative PET leads to watch and wait [with surgery if growth observed on serial CT (tumour resectable) or palliative care if growth observed on serial CT (tumour unresectable)].<sup>24</sup> Positive PET has thoracotomy (80%) or biopsy (20%) (positive biopsies have thoracotomy, otherwise watch and wait); negative PET leads to watch and wait.<sup>21</sup>

g CT then QECT: positive CT has QECT; positive QECT has thoracotomy (80%) or biopsy (20%) [positive biopsies have thoracotomy, otherwise (serial chest X-ray) watch and wait]. Negative QECT watch and wait; negative CT leads to watch and wait. CT then QECT then PET adds PET for positive QECT [positive PET leads to same outcomes as for positive QECT; negative PET leads to thoracotomy or watchful waiting (proportions not clear)].

h CT then CT-guided biopsy: positive CT has CT-guided needle biopsy; positive biopsy has surgery, positive biopsy has CT follow-up; negative CT leads to CT follow-up. CT then PET then CT-guided biopsy adds PET for patients with positive CT [positive PET leads to CT-guided biopsy (80%) or surgery (20%) (biopsy outcomes same as above); negative CT leads to CT follow-up].

TABLE 37 Quality assessment of included studies

Item	Gambhir <i>et al.</i> <sup>22</sup>	Dietlein <i>et al.</i> <sup>24</sup>	Keith <i>et al.</i> <sup>21</sup>	Comber <i>et al.</i> <sup>26</sup>	Gould <i>et al.</i> <sup>20</sup>	Gugiatti <i>et al.</i> <sup>25</sup>	Tsushima and Endo <sup>27</sup>	Lejeune <i>et al.</i> <sup>28</sup>	Deppen <i>et al.</i> <sup>55</sup>
1. Is there a clear statement of the decision problem?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2. Is the comparator routinely used in clinical practice?	No	No	No	No	Yes	Yes	No	Yes	Yes
3. Is the perspective of the model clearly stated?	No	Yes	No	No	Yes	Yes	No	Yes	No
4. Is the study type appropriate? <sup>a</sup>	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
5. Is the modelling methodology appropriate? <sup>a</sup>	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
6. (a) Is the model structure described?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
(b) Does it reflect the disease process?	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
7. Are assumptions about model structure listed and justified?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
8. Are the data inputs for the model described and justified?	Yes	Yes	Yes	Yes	Yes	No <sup>b</sup>	Yes	Yes	Yes
9. Is the effectiveness of the intervention (diagnostic accuracy) established based on a systematic review? <sup>c</sup>	No	No	No	No	Yes	No	No	No	Yes
10. (a) Are health benefits measured in QALYs?	No	No	No	No	Yes	No	No	No	Yes
(b) Are health benefits measured using a standardised and validated generic instrument?	N/A	N/A	N/A	N/A	Yes	N/A	N/A	N/A	No
11. (a) Are resource inputs described and justified?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
(b) Are resources valued appropriately?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
12. Have the costs and outcomes been discounted?	No	Yes <sup>d</sup>	No	No	Yes	No	No	Yes	No
13. Has an incremental analysis of costs and consequences of alternatives been performed?	Yes <sup>e</sup>	Yes <sup>e</sup>	Yes <sup>e</sup>	Yes <sup>e</sup>	Yes	No	Yes <sup>e</sup>	Yes <sup>e</sup>	Yes <sup>e</sup>

TABLE 37 Quality assessment of included studies (*continued*)

Item	Gambhir <i>et al.</i> <sup>22</sup>	Dietlein <i>et al.</i> <sup>24</sup>	Keith <i>et al.</i> <sup>21</sup>	Comber <i>et al.</i> <sup>26</sup>	Gould <i>et al.</i> <sup>20</sup>	Gugiatti <i>et al.</i> <sup>25</sup>	Tsushima and Endo <sup>27</sup>	Lejeune <i>et al.</i> <sup>28</sup>	Deppen <i>et al.</i> <sup>55</sup>
14. Has uncertainty been adequately assessed?	Yes <sup>f</sup>	Yes <sup>f</sup>	No	No	Yes <sup>f</sup>	No	No	No	No

N/A, not applicable.

a Appropriate to the stated aim of the study and current practice at the time. Not necessarily appropriate to current NHS decision-making (*Chapter 5, Quality assessment*).

b There is insufficient detail on where diagnostic accuracy figures come from and how/if they were pooled.

c By systematic review, we mean the following: has a systematic search been conducted (with search strategies documented)?; have inclusion/exclusion criteria been reported?; has an appropriate method of synthesis been conducted and fully reported?

d Study reports using a discount rate of 5%. It appears that this has been used to discount future costs for palliative care only. There does not appear to have been any discounting of outcomes.

e Not a fully incremental analysis. All incremental costs and outcomes are reported against a common baseline, rather than incremental to the next best option.

f The majority of the studies have reported multiple deterministic sensitivity analyses (including some two-way sensitivity analyses varying prevalence of malignancy against diagnostic accuracy and cost parameters). Only Gould *et al.*<sup>20</sup> conducted a probabilistic sensitivity analysis, and report very little detail of the methodology, conduct or outputs.

TABLE 38 Sensitivity analysis reported in included papers

Form of analysis	Type of input	Parameter included in sensitivity analysis	Gambhir <i>et al.</i> <sup>22</sup>	Dietlein <i>et al.</i> <sup>24</sup>	Keith <i>et al.</i> <sup>21</sup>	Comber <i>et al.</i> <sup>26</sup>	Gould <i>et al.</i> <sup>20</sup>	Gugiatti <i>et al.</i> <sup>25</sup>	Tsushima and Endo <sup>27</sup>	Lejeune <i>et al.</i> <sup>28</sup>	Deppen <i>et al.</i> <sup>55</sup>	
Deterministic sensitivity analysis	Disease/node characteristics	Probability of malignancy <sup>a</sup>	✓	✓	✓	✓	✓	✓	✓	✓	✓	
		Nodule size	✓				✓					
		Growth of benign nodule					✓					
	Test accuracy	CT sensitivity					✓	✓			✓	
		CT specificity	✓				✓	✓			✓	
		PET sensitivity	✓	✓ <sup>b</sup>			✓	✓			✓	✓
		PET specificity	✓	✓ <sup>b</sup>			✓	✓			✓	✓
		Biopsy sensitivity	✓ <sup>c</sup>	✓			✓				✓	✓
		Biopsy specificity	✓ <sup>c</sup>				✓				✓	✓
		Costs	CT cost					✓				
	PET cost	✓	✓	✓	✓ <sup>d</sup>	✓				✓		
	Surgery cost	✓				✓						
	Biopsy cost	✓ <sup>c</sup>				✓						
	Cost of palliative care			✓								
	Morbidity/mortality	Surgery mortality rate			✓		✓					
		Procedure-related morbidity					✓					
		Adjustment of life expectancy by hospital stay									✓	
	Other	Mix strategy	✓ <sup>e</sup>									
		Risk patients <sup>a</sup>	✓ <sup>f</sup>	✓ <sup>g</sup>							✓ <sup>h</sup>	

Form of analysis	Type of input	Parameter included in sensitivity analysis	Gambhir <i>et al.</i> <sup>22</sup>	Dietlein <i>et al.</i> <sup>24</sup>	Keith <i>et al.</i> <sup>21</sup>	Comber <i>et al.</i> <sup>26</sup>	Gould <i>et al.</i> <sup>20</sup>	Gugiatti <i>et al.</i> <sup>25</sup>	Tsushima and Endo <sup>27</sup>	Lejeune <i>et al.</i> <sup>28</sup>	Deppen <i>et al.</i> <sup>55</sup>
Probabilistic sensitivity analysis			X	X	X	X	✓	X	X	X	X
<p>a Often in primary analysis – addressing robustness of results for different groups of patients/risk factors (age, SPN diameter, smoking history), whereas other analyses consider robustness for the same patient group.</p> <p>b Tested simultaneously.</p> <p>c Tested simultaneously in one analysis of biopsy type (transthoracic needle aspiration biopsy vs. video-assisted thoracoscopic surgery).</p> <p>d Expressed as a ratio versus surgery cost.</p> <p>e In base case, 80% of patients with CT/PET positive nodule proceed to immediate surgery. Remaining 20% undergo biopsy. Positive biopsies undergo surgery; negative biopsies assigned to wait and watch.</p> <p>f Presented in primary analysis for five different patient characteristic sets (age, SPN diameter, current smoking status) with pre-test likelihood of malignancy varying from 0.001 to 0.80.</p> <p>g The paper does not state what they mean by risk patients nor do they discuss the result of this sensitivity analysis.</p> <p>h Presented for four different patient characteristic sets (age, SPN diameter, current smoking status), with pre-test likelihood of malignancy varying from 0.03 to 0.87.</p>											

TABLE 39 Cost-effectiveness of individual tests and diagnostic strategies for SPNs

Study (outcome) [currency]	Prevalence (percentage expressed as a decimal)	Nodule size (cm)	Single tests							Combination strategies			
			Observation/ watch and wait	CT	PET	Biopsy	Immediate surgery	QECT	QECT + <sup>18</sup> F-FDG- PET/CT	CT + PET	CT + biopsy	CT + PET + biopsy	CT-FNA
<b>Gambhir et al.<sup>22</sup> (ICER) [US\$]</b>													
64-year-old, male smoker	0.83	2.5	Baseline	3266						4273			
35-year-old, male non-smoker	0.001	1.0	Baseline	-63,606						-205,172			
50-year-old, male smoker	0.34	2.0	Baseline	11,662						7549			
75-year-old, male non-smoker	0.23	2.0	Baseline	94,449						39,457			
75-year-old, male smoker	0.80	2.0	Baseline	6567						7993			
<b>Dietlein et al.<sup>24</sup> (ICER) [Euro]</b>													
62-year-old, male	0.65	< 3	Baseline			3218	6120	4210					
62-year-old, male	0.65	< 3	4210			-6912	3343	Baseline					
<b>Keith et al.<sup>21</sup> (ICAR) [Euro]</b>													
67-year-old, people	0.54	NR	Baseline	10,231	6433								
67-year-old, people	0.54	NR	Baseline	8788	6277								
<b>Comber et al.<sup>26</sup> (ICAR) [AUS\$]</b>													
NR	0.54	NR	Baseline	16,847	12,300					12,636		12,059	

Study (outcome) [currency]	Prevalence (percentage expressed as a decimal)	Single tests					Combination strategies						
		Nodule size (cm)	Observation/ watch and wait	CT	PET	Biopsy	Immediate surgery	QECT	CT + PET	QECT + <sup>18</sup> F-FDG- PET/CT	CT + PET + biopsy	CT + PET + biopsy	CT-FNA
<i>Tsushima and Endo</i> <sup>27</sup> (ICAR) [Japanese Yen]													
NR	0.10	1-4		Baseline							-14,037	-15,805	-15,013
<i>Lejeune et al.</i> <sup>28</sup> (ICER) [Euro]													
65-year-old, male current smoker	0.43	2	Baseline		4790						3022		
<i>Deppen et al.</i> <sup>55</sup> (ICER) [US\$]													
60-year-old, male smoker	0.5	1.5-2			Baseline	7333					-118,350		9533
60-year-old, male smoker	0.65	1.5-2			Baseline	4602					43,578		3998
CT-FNA, computerised tomography-guided fine-needle aspiration; ICAR, incremental cost-accuracy ratio; NR, not reported.													

TABLE 40 The ICERs for non-dominated strategies at different pre-test probabilities of malignancy reported by Gould *et al.*<sup>20</sup>

Strategy	ICER (US\$)
<b>Low probability of malignancy (0.26)</b>	
1. Watchful waiting	Baseline
7. Do CT: if results indeterminate, do biopsy; if results benign, watch and wait	10,935
12. Do CT: if results indeterminate, do <sup>18</sup> F-FDG-PET/CT; if <sup>18</sup> F-FDG-PET/CT results positive, do surgery; if <sup>18</sup> F-FDG-PET/CT results negative, do biopsy; if CT results benign, watch and wait	20,445
15. Do CT: if results indeterminate, do <sup>18</sup> F-FDG-PET/CT; if <sup>18</sup> F-FDG-PET/CT results positive, do surgery; if <sup>18</sup> F-FDG-PET/CT results negative, do biopsy; if CT results benign, do biopsy	45,838
35. Do CT and <sup>18</sup> F-FDG-PET/CT: if CT results indeterminate and <sup>18</sup> F-FDG-PET/CT results positive, do surgery; if CT results indeterminate and <sup>18</sup> F-FDG-PET/CT results negative, do biopsy; if CT results benign and <sup>18</sup> F-FDG-PET/CT results positive, do biopsy	297,212
<b>Intermediate probability of malignancy (0.55)</b>	
1. Watchful waiting	Baseline
7. Do CT: if results indeterminate, do biopsy; if results benign, watch and wait	7625
3. Do CT-guided biopsy	14,981
6. Do CT: if results indeterminate, do surgery; if results benign, do biopsy	17,649
15. Do CT: if results indeterminate, do <sup>18</sup> F-FDG-PET/CT; if <sup>18</sup> F-FDG-PET/CT results positive, do surgery; if <sup>18</sup> F-FDG-PET/CT results negative, do biopsy; if CT results benign, do biopsy	229,260
31. Do <sup>18</sup> F-FDG-PET/CT: if <sup>18</sup> F-FDG-PET/CT results negative, do CT; if CT results indeterminate, do biopsy; if CT results benign, watch and wait; if <sup>18</sup> F-FDG-PET/CT results positive, do surgery	288,910
<b>High probability of malignancy (0.55)</b>	
1. Watchful waiting	Baseline
5. Do CT: if results indeterminate, do surgery; if results benign, watch and wait	6515
19. Do CT: if CT results indeterminate, do surgery; if CT results benign, do <sup>18</sup> F-FDG-PET/CT; if <sup>18</sup> F-FDG-PET/CT results positive, do biopsy; if <sup>18</sup> F-FDG-PET/CT results negative, watch and wait	16,261
18. Do CT: if CT results indeterminate, do surgery; if CT results benign, do <sup>18</sup> F-FDG-PET/CT; if <sup>18</sup> F-FDG-PET/CT results positive, do surgery; if <sup>18</sup> F-FDG-PET/CT results negative, watch and wait	50,839
17. Do CT: if CT results indeterminate, do surgery; if CT results benign, do <sup>18</sup> F-FDG-PET/CT; if <sup>18</sup> F-FDG-PET/CT results positive, do surgery; if <sup>18</sup> F-FDG-PET/CT results negative, biopsy	67,568
Adapted from table 1 of Gould <i>et al.</i> <sup>20</sup> Modified with permission from <i>Annals of Internal Medicine</i> . <sup>20</sup>	

## Appendix 10 List of centres and principal investigators

Centre	Hospital trust	PI
Aberdeen Royal Infirmary	NHS Grampian	Dr Lesley Gomersall
Churchill Hospital (Oxford)	Oxford University Hospitals NHS Foundation Trust	Professor Fergus V Gleeson
Conquest Hospital, Hastings	East Sussex Healthcare NHS Trust	Dr Osei Kankam
Glasgow Royal Infirmary	NHS Greater Glasgow and Clyde	Dr Sai Han
Leicester Royal Infirmary	University Hospitals of Leicester NHS Trust	Dr Jonathan Bennett
Nottingham City Hospital	Nottingham University Hospitals NHS Trust	Professor David Baldwin
Royal Papworth Hospital (Cambridge)	Royal Papworth Hospital NHS Foundation Trust	Dr Nagmi R Qureshi
Royal Infirmary of Edinburgh	NHS Lothian	Dr Kristopher Skwarski
Royal Sussex County Hospital (Brighton)	Brighton and Sussex University Hospitals NHS Trust	Dr Sabina Dizdarevic
Southampton General Hospital	University Hospital Southampton NHS Foundation Trust	Dr Anindo Banerjee
St. James's University Hospital (Leeds)	Leeds Teaching Hospitals NHS Trust	Dr Matthew E Callister
University College Hospital, London	University College London Hospitals NHS Foundation Trust	Professor Ashley M Groves
University Hospital of South Manchester (Wythenshawe)	Manchester University NHS Foundation Trust	Dr Philip Crosbie
Weston General Hospital (Weston-super-Mare)	University Hospitals Bristol and Weston NHS Foundation Trust	Dr John O'Brien
Worcestershire Royal Hospital	Worcestershire Acute Hospitals NHS Trust	Dr Steve O'Hickey
Worthing Hospital	Western Sussex Hospitals NHS Foundation Trust	Dr Nick Adams



## Appendix 11 Main results: additional tables

### Recruitment by site

Table 41 shows the study recruitment by site, as well as the breakdown of which sites provided data for the final analysis set of 312 participants.

### Diagnostic performance of the imaging techniques

#### Alternative prevalence situations

Tables 42 and 43 show the change in diagnostic performance (for the pre-defined thresholds) as the prevalence changes to either 50% or 70%. Performance across the models becomes more similar as the prevalence increases to 70%, and the difference widens in lower prevalence situations, where the difference in specificity becomes more important.

#### Subgroup analysis by size

Tables 44–46 show how the diagnostic performance of the pre-defined thresholds varies as the size of the nodule is altered. As the nodules get larger, the performance of most of the models increases.

TABLE 41 Trial recruitment and data availability summary by site

Site	Date opened	Total recruited (n)	Usable PET/CT and DCE-CT data, n (%)	Two-year outcome status for the nodule, n (%)
Royal Papworth Hospital	17 December 2012	74	71 (96)	71 (100)
Southampton	8 January 2013	21	20 (95)	19 (95)
Glasgow	24 January 2013	41	40 (98)	40 (100)
UCLH	3 May 2013	23	18 (78)	18 (100)
Aberdeen	4 June 2013	38	35 (92)	34 (97)
Brighton	25 September 2013	22	19 (86)	19 (100)
Leeds	11 November 2013	54	43 (80)	42 (98)
Manchester	25 February 2014	17	10 (59)	10 (100)
Oxford	31 March 2014	25	20 (80)	19 (95)
Worcester	23 April 2014	17	9 (53)	9 (100)
Worthing	15 August 2014	3	0 (0)	N/A
Weston	19 February 2015	2	2 (100)	2 (100)
Hastings	19 April 2015	3	3 (100)	3 (100)
Nottingham	28 April 2015	24	15 (63)	15 (100)
Leicester	29 July 2015	12	10 (83)	9 (90)
Edinburgh	2 September 2015	4	2 (50)	2 (100)
Total		380	317 (83)	312 (98)

N/A, not applicable; UCLH, University College London Hospital.

TABLE 42 Diagnostic performance of each imaging technique: pre-defined thresholds with a 50% prevalence (N = 312)

Objective	Imaging technique	Sensitivity, n/N (%) (95% CI)	Specificity, n/N (%) (95% CI)	NPV, n/N (%) (95% CI)	PPV, n/N (%) (95% CI)	ODA, n/N (%) (95% CI)
Primary objective 1	DCE-CT (maximum enhancement $\geq$ 15 HU)	151/156 (96.8) (92.7% to 98.6%)	32/156 (20.5) (14.9% to 27.5%)	32/37 (86.5) (72.0% to 94.1%)	151/275 (54.9) (49.0% to 60.7%)	183/312 (58.8) (53.1% to 64.0%)
	DCE-CT (maximum enhancement $\geq$ 20 HU)	149/156 (95.5) (91.0% to 97.8%)	46/156 (29.5) (22.9% to 37.1%)	46/53 (86.8) (75.2% to 93.5%)	149/259 (57.5) (51.4% to 63.4%)	195/312 (62.6) (57.0% to 67.7%)
	PET/CT (based on PET and CT grading)	114/156 (73.1) (65.6% to 79.4%)	128/156 (82.1) (75.3% to 87.3%)	128/170 (75.3) (68.3% to 81.2%)	114/142 (80.3) (73.0% to 86.0%)	241/312 (77.2) (72.3% to 81.6%)
Secondary objective 1	PET/CT (based on PET grading alone)	118/156 (75.6) (68.3% to 81.7%)	132/156 (84.6) (78.1% to 89.4%)	132/170 (77.7) (70.8% to 83.3%)	118/142 (83.1) (76.1% to 88.4%)	249/312 (79.8) (75.0% to 83.9%)
	PET/CT (based on a SUVmax of $\geq$ 2.5)	119/156 (76.3) (69.0% to 82.3%)	127/156 (81.4) (74.6% to 86.7%)	127/164 (77.4) (70.5% to 83.2%)	119/148 (80.4) (73.3% to 86.0%)	246/312 (78.9) (74.0% to 83.0%)
Secondary objective 2	Combination of DCE-CT and PET/CT	110/156 (70.5) (62.9% to 77.1%)	130/156 (83.3) (76.7% to 88.4%)	130/177 (73.5) (66.5% to 79.4%)	110/135 (81.5) (74.1% to 87.1%)	240/312 (76.9) (71.9% to 81.3%)

TABLE 43 Diagnostic performance of each imaging technique: pre-defined thresholds with a 70% prevalence (N = 312)

Objective	Imaging technique	Sensitivity, n/N (%) (95% CI)	Specificity, n/N (%) (95% CI)	NPV, n/N (%) (95% CI)	PPV, n/N (%) (95% CI)	ODA, n/N (%) (95% CI)
Primary objective 1	DCE-CT (maximum enhancement $\geq$ 15 HU)	212/218 (97.3) (94.1% to 98.7%)	19/94 (20.2) (13.3% to 29.4%)	19/26 (73.1) (53.9% to 86.3%)	212/286 (74.1) (68.8% to 78.9%)	231/312 (74.0) (68.9% to 78.6%)
	DCE-CT (maximum enhancement $\geq$ 20 HU)	208/218 (95.4) (91.8% to 97.5%)	28/94 (29.8) (21.5% to 39.7%)	28/38 (73.7) (58.0% to 85.0%)	208/274 (75.9) (70.5% to 80.6%)	236/312 (75.6) (70.6% to 80.1%)
	PET/CT (based on PET and CT grading)	159/218 (72.9) (66.7% to 78.4%)	77/94 (81.9) (72.9% to 88.4%)	77/136 (56.6) (48.2% to 64.7%)	159/176 (90.3) (85.1% to 93.9%)	236/312 (75.6) (70.6% to 80.1%)
Secondary objective 1	PET/CT (based on PET grading alone)	165/218 (75.7) (69.6% to 80.9%)	78/94 (83.0) (74.1% to 89.2%)	79/133 (59.4) (50.9% to 67.4%)	165/179 (92.2) (87.3% to 95.3%)	244/312 (78.2) (73.3% to 82.4%)
	PET/CT (based on a SUVmax of $\geq$ 2.5)	167/218 (76.6) (70.6% to 81.7%)	76/94 (80.9) (71.8% to 87.5%)	76/128 (59.4) (50.7% to 67.5%)	167/184 (90.8) (85.7% to 94.2%)	243/312 (77.9) (73.0% to 82.1%)
Secondary objective 2	Combination of DCE-CT and PET/CT	153/218 (70.2) (63.8% to 75.9%)	78/94 (83.0) (74.1% to 89.2%)	78/143 (54.6) (46.4% to 62.5%)	153/169 (90.5) (85.2% to 94.1%)	231/312 (74.0) (68.9% to 78.6%)

TABLE 44 Diagnostic performance of each imaging technique: pre-defined thresholds in subset of patients with a SPN of 8–15 mm on initial CT scan (N = 167)

Imaging technique	Sensitivity, n/N (%) (95% CI)	Specificity, n/N (%) (95% CI)	NPV, n/N (%) (95% CI)	PPV, n/N (%) (95% CI)	ODA, n/N (%) (95% CI)
DCE-CT (maximum enhancement $\geq 15$ HU)	87/89 (97.8) (92.2% to 99.4%)	15/78 (19.2) (12.0% to 29.3%)	15/17 (88.2) (65.7% to 96.7%)	87/150 (58.0) (50.0% to 65.6%)	102/167 (61.1) (53.5% to 68.1%)
DCE-CT (maximum enhancement $\geq 20$ HU)	87/89 (97.8) (92.2% to 99.4%)	22/78 (28.2) (19.4% to 39.0%)	22/24 (91.7) (74.2% to 97.7%)	87/143 (60.1) (52.6% to 68.5%)	109/167 (65.3) (57.8% to 72.1%)
PET/CT (based on PET and CT grading)	57/89 (64.0) (53.7% to 73.2%)	69/78 (88.5) (79.5% to 93.8%)	69/101 (68.3) (58.7% to 76.6%)	57/66 (86.4) (76.1% to 92.7%)	126/167 (75.5) (68.4% to 81.4%)
PET/CT (based on PET grading alone)	57/89 (64.0) (53.7% to 73.2%)	68/78 (87.2) (78.0% to 92.9%)	68/100 (68.0) (58.3% to 76.3%)	57/67 (85.1) (74.7% to 91.7%)	125/167 (74.9) (67.8% to 80.8%)
PET/CT (based on a SUVmax of $\geq 2.5$ )	57/89 (64.0) (53.7% to 73.2%)	66/76 (86.8) (77.5% to 92.7%)	66/98 (67.4) (57.6% to 75.8%)	57/67 (85.1) (74.7% to 91.7%)	123/165 (74.6) (67.4% to 80.6%)
Combination of DCE-CT and PET/CT	57/89 (64.0) (53.7% to 73.2%)	69/78 (88.5) (79.5% to 93.8%)	69/101 (68.3) (58.7% to 76.6%)	57/66 (86.4) (76.1% to 92.7%)	126/167 (75.5) (68.4% to 81.4%)

TABLE 45 Diagnostic performance of each imaging technique: pre-defined thresholds in subset of patients with a SPN of > 15 mm to  $\leq 20$  mm on initial CT scan (N = 74)

Imaging technique	Sensitivity, n/N (%) (95% CI)	Specificity, n/N (%) (95% CI)	NPV, n/N (%) (95% CI)	PPV, n/N (%) (95% CI)	ODA, n/N (%) (95% CI)
DCE-CT (maximum enhancement $\geq 15$ HU)	46/47 (97.9) (88.9% to 99.6%)	6/27 (22.2) (10.6% to 40.8%)	6/7 (85.7) (48.7% to 97.4%)	46/67 (68.7) (56.8% to 78.5%)	52/74 (70.3) (59.1% to 79.5%)
DCE-CT (maximum enhancement $\geq 20$ HU)	44/47 (93.6) (82.8% to 97.8%)	9/27 (33.3) (18.6% to 52.2%)	9/12 (75.0) (46.8% to 91.1%)	44/62 (71.0) (58.7% to 80.8%)	53/74 (71.6) (60.5% to 80.6%)
PET/CT (based on PET and CT grading)	36/47 (76.6) (62.8% to 86.4%)	18/27 (66.7) (47.8% to 81.4%)	18/29 (62.1) (44.0% to 77.3%)	36/45 (80.0) (66.2% to 89.1%)	54/74 (73.0) (61.9% to 81.8%)
PET/CT (based on PET grading alone)	37/47 (78.7) (65.1% to 88.0%)	21/27 (77.8) (59.2% to 89.4%)	21/31 (67.7) (50.1% to 81.4%)	37/43 (86.1) (72.7% to 93.4%)	58/74 (78.4) (67.7% to 86.2%)
PET/CT (based on a SUVmax of $\geq 2.5$ )	38/47 (80.9) (67.5% to 89.6%)	20/27 (74.1) (55.3% to 86.8%)	20/29 (69.0) (50.8% to 82.7%)	38/45 (84.4) (71.2% to 92.3%)	58/74 (78.4) (67.7% to 86.2%)
Combination of DCE-CT and PET/CT	34/47 (72.3) (58.2% to 83.1%)	18/27 (66.7) (47.8% to 81.4%)	18/31 (58.1) (40.8% to 73.6%)	34/43 (79.1) (64.8% to 88.6%)	52/74 (70.3) (59.1% to 79.5%)

TABLE 46 Diagnostic performance of each imaging technique: pre-defined thresholds in subset of patients with a SPN of &gt; 20 mm on initial CT scan (N = 65)

Imaging technique	Sensitivity, n/N (%) (95% CI)	Specificity, n/N (%) (95% CI)	NPV, n/N (%) (95% CI)	PPV, n/N (%) (95% CI)	ODA, n/N (%) (95% CI)
DCE-CT (maximum enhancement $\geq$ 15 HU)	49/52 (94.2) (84.4% to 98.0%)	4/13 (30.8) (12.7% to 57.6%)	4/7 (57.1) (25.1% to 84.2%)	49/58 (84.5) (73.1% to 91.6%)	53/65 (81.5) (70.5% to 89.1%)
DCE-CT (maximum enhancement $\geq$ 20 HU)	49/52 (94.2) (84.4% to 98.0%)	4/13 (30.8) (12.7% to 57.6%)	4/7 (57.1) (25.1% to 84.2%)	49/58 (84.5) (73.1% to 91.6%)	53/65 (81.5) (70.5% to 89.1%)
PET/CT (based on PET and CT grading)	45/52 (86.5) (74.7% to 93.3%)	9/13 (69.2) (42.4% to 87.3%)	9/16 (56.3) (33.2% to 76.9%)	45/49 (91.8) (80.8% to 96.8%)	54/65 (83.1) (72.2% to 90.3%)
PET/CT (based on PET grading alone)	49/52 (94.2) (84.4% to 98.0%)	10/13 (76.9) (49.7% to 91.8%)	10/13 (76.9) (49.7% to 91.8%)	49/52 (94.2) (84.4% to 98.0%)	59/65 (90.8) (81.3% to 95.7%)
PET/CT (based on a SUVmax of $\geq$ 2.5)	50/52 (96.2) (87.0% to 98.9%)	8/13 (61.5) (35.5% to 82.3%)	8/10 (80.0) (49.0% to 94.3%)	50/55 (90.9) (80.4% to 96.1%)	58/65 (89.2) (79.4% to 94.7%)
Combination of DCE-CT and PET/CT	43/52 (82.7) (70.3% to 90.6%)	11/13 (84.6) (57.8% to 95.7%)	11/20 (55.0) (34.2% to 74.2%)	43/45 (95.6) (85.2% to 98.8%)	54/65 (83.1) (72.2% to 90.3%)

## Adverse events

Seventeen adverse events were experienced by 10 study participants, with four related adverse events (possibly, probably or definitely related) experienced by four study participants. The adverse events are documented in *Table 47*. One serious adverse event was reported during the course of the study: a post-biopsy haemoptysis, necessitating prolonged hospitalisation.

TABLE 47 Number of participants with an adverse event who reported each type

Adverse event	Severity	Total (N = 10), n (%)
Itch and rash	Mild	4 (40)
Vomiting	Mild	2 (20)
Diarrhoea	Mild	1 (10)
Neck, back and shoulder pain	Mild	1 (10)
Night-time leg cramps	Mild	1 (10)
Dizziness	Mild	1 (10)
Allergic reaction	Mild	1 (10)
Skin reaction (feet)	Moderate	1 (10)



## Appendix 12 Applying likelihood ratios to update prevalence of disease calculations

In this appendix, a worked example of how the posterior probability of malignancy was defined throughout the model is presented, using an example with notation in *Table 48*.

In addition, the estimation of the probability of a determinate result is explained.

Based on the notation in *Table 48*, worked examples are provided, based on the imaging test results of  $^{18}\text{F}$ -FDG-PET/CT (*Table 49*).

First, the prevalence of malignancy is determined by the total number of malignant and benign cases:

$$\text{Prevalence} = \frac{\text{Malignant}}{\text{Malignant} + \text{Benign}} = \frac{g}{g + h} = \frac{189}{312} = 0.6057. \quad (11)$$

The sensitivity and specificity for a determinate test result are based on the diagnostic tests yield in patients with malignant and benign tumours:

$$\text{Sensitivity (diagnostic)} = \frac{a + c}{g} = \frac{4 + 81}{189} = 0.4497. \quad (12)$$

$$\text{Specificity (diagnostic)} = \frac{f}{h} = \frac{87}{123} = 0.707. \quad (13)$$

TABLE 48 Notation of diagnostic results of  $^{18}\text{F}$ -FDG-PET/CT according to malignancy status

Test outcome	Malignancy status	
	Malignant	Benign
Cancer	<i>a</i>	<i>b</i>
Non-cancer	<i>c</i>	<i>d</i>
Indeterminate	<i>e</i>	<i>f</i>
Total cases	<i>g</i>	<i>h</i>

TABLE 49 Diagnostic results of  $^{18}\text{F}$ -FDG-PET/CT according to malignancy status

Test outcome	Malignancy status		Total cases
	Malignant	Benign	
Non-cancer	4	27	–
Undetermined	104	87	–
Lung cancer	81	9	–
Total cases	189	123	312

Knowing the sensitivity and specificity for a determinate result, the posterior prevalence of malignancy in the population with diagnostic test results is determined by applying the PLR to the pre-test prevalence, expressed as an OR:

$$PLR = \frac{Sensitivity}{1 - specificity} = \frac{0.4497}{0.293} = 1.5348. \quad (14)$$

$$OR_{(prior\ prevalence)} = \frac{Prevalence}{1 - prevalence} = \frac{0.6057}{0.3943} = 1.536. \quad (15)$$

$$OR_{(posterior\ prevalence)} = \frac{Prevalence}{1 - prevalence} \times \frac{Sensitivity}{1 - specificity} = 1.54 \times 1.536 = 2.36. \quad (16)$$

The final step was to transform these ORs into a probability ( $p$ ), to estimate the posterior prevalence of malignancy, following a diagnostic test result:

$$p_{posterior\ malignancy} = \frac{OR_{(posterior\ prevalence)}}{OR_{(posterior\ prevalence)} + 1} = 0.702. \quad (17)$$

Consequently, following a  $^{18}\text{F}$ -FDG-PET/CT determinate result, the posterior prevalence of malignancy was 70.2%. This process was repeated for all diagnostic tests throughout the model to estimate the posterior probability of malignancy, conditional on a diagnostic test's result.

Furthermore, the probability of a determinate test result was conditional on prevalence and was as follows:

$$p_{Determinate} = (Sensitivity \times prevalence) + [(1 - prevalence) \times (1 - specificity)] = (0.4497 \times 0.6057) + (0.293 \times 0.3943) = 0.3879. \quad (18)$$

The same method was applied to estimate the probability of a positive test result, given that the test had originally yielded a determinate result.

## Appendix 13 Parameter tables

TABLE 50 Clinical parameters used in the model

Description	Mean	Distribution (parameter values)	Source
Prevalence	0.606	Beta (189, 123)	SputNIK trial
<sup>18</sup> F-FDG-PET/CT sensitivity for determinate results	0.450	Dirichlet (4, 104, 81, 27, 87, 9) <sup>a</sup>	SputNIK trial
<sup>18</sup> F-FDG-PET/CT specificity for determinate results	0.707	Dirichlet (4, 104, 81, 27, 87, 9) <sup>a</sup>	SputNIK trial
<sup>18</sup> F-FDG-PET/CT sensitivity provided that a determinate diagnosis has been obtained	0.953	Dirichlet (4, 104, 81, 27, 87, 9) <sup>a</sup>	SputNIK trial
<sup>18</sup> F-FDG-PET/CT specificity provided that a determinate diagnosis has been obtained	0.750	Dirichlet (4, 104, 81, 27, 87, 9) <sup>a</sup>	SputNIK trial
DCE-CT sensitivity for determinate results	0.258	Dirichlet (4, 138, 44, 27, 89, 7) <sup>a</sup>	SputNIK trial
DCE-CT specificity for determinate results	0.724	Dirichlet (4, 138, 44, 27, 89, 7) <sup>a</sup>	SputNIK trial
DCE-CT sensitivity provided that a determinate diagnosis has been obtained	0.917	Dirichlet (4, 138, 44, 27, 89, 7) <sup>a</sup>	SputNIK trial
DCE-CT specificity provided that a determinate diagnosis has been obtained	0.794	Dirichlet (4, 138, 44, 27, 89, 7) <sup>a</sup>	SputNIK trial
Biopsy diagnostic yield	0.934	Gaussian (0.934, 0.015)	Han <i>et al.</i> <sup>66</sup>
Biopsy sensitivity	0.912	NA	Appendix 14
Biopsy specificity	0.955	NA	Appendix 14
Biopsy sensitivity and specificity correlation	-2.430	NA	Appendix 14
Histopathology on surgical sample sensitivity	1	NA	Assumption
Histopathology on surgical sample specificity	1	NA	Assumption
Growing nodule status sensitivity	0.457	Dirichlet (21, 25, 27, 64)	SputNIK trial
Growing nodule status specificity	0.703	Dirichlet (21, 25, 27, 64)	SputNIK trial
NA, not available.			
a In the absence of correlation between sensitivity and specificity, Dirichlet distributions were used. These are based on the counts of diagnostic test results from the trial. Measures of diagnostic accuracy were then calculated based on the sampled counts from the Dirichlet distributions.			

TABLE 51 Estimated proportions for procedure-related non-fatal complications and mortality

Description	Mean	95% CI	Source
Biopsy non-fatal complications	0.030	0.018 to 0.048	Han <i>et al.</i> <sup>66</sup>
Biopsy operative mortality	0.004	0.003 to 0.006	Weiner <i>et al.</i> <sup>67</sup>
Lobectomy major non-fatal complications	0.064	NA	Deppen <i>et al.</i> <sup>55</sup>
Lobectomy operative mortality	0.022	NA	Deppen <i>et al.</i> <sup>55</sup>
Wedge resection major non-fatal complications	0.046	NA	Deppen <i>et al.</i> <sup>55</sup>
Wedge resection operative mortality	0.012	NA	Deppen <i>et al.</i> <sup>55</sup>
Radiotherapy non-fatal complications (moderate grade)	0.073	0.061 to 0.082	Zhao <i>et al.</i> <sup>68</sup>
Radiotherapy non-fatal complications (severe grade)	0.022	NA	Zhao <i>et al.</i> <sup>68</sup>
NA, not available.			

TABLE 52 Distribution of management options following initial imaging test results (from trial)

Description	Estimated proportion	Distribution (parameter values)
<b>Following positive imaging test results</b>		
Probability of biopsy	0.630	Dirichlet (51, 27, 3)
Probability of surgery	0.333	
Probability of radiotherapy	0.037	
<b>Following negative imaging test results</b>		
Probability of biopsy	0.133	Dirichlet (4, 2, 24)
Probability of surgery	0.067	
Probability of follow-up CT	0.800	
<b>Following indeterminate imaging test results</b>		
Probability of biopsy	0.335	Dirichlet (71, 41, 100)
Probability of surgery	0.193	
Probability of follow-up CT	0.472	
<b>Following positive imaging test results and indeterminate first biopsy</b>		
Probability of second biopsy	0.222	Dirichlet (2, 6, 1)
Probability of surgery or SABR	0.667	
Probability of no treatment	0.111	
<b>Following negative imaging test results and indeterminate first biopsy</b>		
Probability of second biopsy	0.000	Dirichlet (0, 1, 0)
Probability of surgery or SABR	1.000	
Probability of no treatment	0.000	
<b>Following indeterminate imaging test results and indeterminate first biopsy</b>		
Probability of second biopsy	0.167	Dirichlet (2, 5, 5)
Probability of surgery or SABR	0.417	
Probability of no treatment	0.417	
<b>After second biopsy</b>		
Probability of surgery	0.000	Dirichlet (0, 0, 4)
Probability of radiotherapy	0.000	
Probability of follow-up CT	1.000	
<b>Following a positive imaging test result during the watch-and-wait strategy</b>		
Probability of biopsy for growing nodules	0.000	Dirichlet (0, 1, 3)
Probability of treatment for growing nodules	0.250	
Probability of no treatment for growing nodules	0.750	
Probability of biopsy for stable nodules	0.000	Dirichlet (0, 3, 1)
Probability of treatment for stable nodules	0.750	
Probability of no treatment for stable nodules	0.250	

TABLE 52 Distribution of management options following initial imaging test results (from trial) (continued)

Description	Estimated proportion	Distribution (parameter values)
<b>Following a negative imaging test result during the watch-and-wait strategy</b>		
Probability of biopsy for growing nodules	0.200	Dirichlet (1, 0, 4)
Probability of treatment for growing nodules	0.000	
Probability of no treatment for growing nodules	0.800	
Probability of biopsy for stable nodules	0.000	Dirichlet (0, 0, 19)
Probability of treatment for stable nodules	0.000	
Probability of no treatment for stable nodules	1.000	
<b>Following an indeterminate imaging test result during the watch-and-wait strategy</b>		
Probability of biopsy for growing nodules	0.122	Dirichlet (5, 11, 25)
Probability of treatment for growing nodules	0.268	
Probability of no treatment for growing nodules	0.610	
Probability of biopsy for stable nodules	0.033	Dirichlet (2, 8, 49)
Probability of treatment for stable nodules	0.136	
Probability of no treatment for stable nodules	0.831	

TABLE 53 Unit costs of imaging tests and treatment procedures

Description	Unit cost (£)	Distribution	HRG code	Reference
<sup>18</sup> F-FDG-PET/CT	610	Gamma (1, 0.001639) <sup>a</sup>	RN01A (IMAGOP)	NHS Reference Costs 2017/18 <sup>65</sup>
DCE-CT	133	Gamma (1, 0.007519) <sup>a</sup>	RD22Z (IMAGOP)	NHS Reference Costs 2017/18 <sup>65</sup>
CT	90	Gamma (1, 0.011111) <sup>a</sup>	RD20A (IMAG)	NHS Reference Costs 2017/18 <sup>65</sup>
Biopsy	660	Gamma (1, 0.001500) <sup>a</sup>	YD03Z (w/a of outpatient procedures, day cases, and non-elective short-stay cases)	NHS Reference Costs 2017/18 <sup>65</sup>
Surgery	5778	Gamma (1, 0.000173) <sup>a</sup>	DZ02K	NHS Reference Costs 2017/18 <sup>65</sup>
Radiotherapy	2024	Gamma (1, 0.000491)	YD01Z	NHS Reference Costs 2017/18 <sup>65</sup>
<b>Complications</b>				
Surgery complications	2599	Gamma (1, 0.000385)		Assumption based on NHS Reference Costs 2017/18 <sup>65</sup>
Biopsy complications	523.6	NA		Assumption based on Weiner <i>et al.</i> <sup>67</sup> and cost of pneumothorax (see Appendix 14)
Pneumothorax	308	Gamma (1, 0.00191)	DZ26G-DZ26L	NHS Reference Costs 2017/18 <sup>65</sup>

continued

TABLE 53 Unit costs of imaging tests and treatment procedures (continued)

Description	Unit cost (£)	Distribution	HRG code	Reference
Radiotherapy complications	336.6	NA		Assumption based on LCA Acute Oncology Clinical Guidelines <sup>70</sup> and Paix <i>et al.</i> <sup>72</sup>
Pneumonitis grade 2	8.87	Gamma (1, 0.112740)		Assumption based on Oncology Clinical Guidelines <sup>70</sup> and the BNF <sup>71</sup>
Pneumonitis grades 3–5	327.73	Gamma (1, 0.003051)		Assumption based on Paix <i>et al.</i> , <sup>72</sup> the BNF <sup>71</sup> , and Echevarria <i>et al.</i> <sup>73</sup>
Oxygen at home for 3 months	305.96	NA		Echevarria <i>et al.</i> <sup>73</sup>

BNF, *British National Formulary*; HRG, Healthcare Resource Group; LCA, London Cancer Alliance; NA, not available; w/a, weighted average.  
 a Cost parameters for which point estimates were used in the base-case analysis.

TABLE 54 Life expectancy according to malignancy and treatment status

Life expectancy	Median (years)	Source
Benign	17.41	Based on UK life tables <sup>77</sup>
Wedge resection	17.41	Assumption
Lobectomy	8.83	Paix <i>et al.</i> <sup>72</sup>
SABR (for malignant nodule)	3.99	Paix <i>et al.</i> <sup>72</sup>
SABR (for benign nodule)	17.41	Assumption
Untreated malignant nodule	1.38	Rosen <i>et al.</i> <sup>79</sup>
Utility	Average	
Benign cases (general population)	0.78 (aged 65–74 years), 0.73 (aged ≥ 75 years)	Kind <i>et al.</i> <sup>80</sup>
Progression-free survival	0.712	Doyle <i>et al.</i> <sup>81</sup>
Untreated	0.421	Assumption based on Doyle <i>et al.</i> <sup>81</sup>
Utility decrement	Per event	
Biopsy complications	0.132 (for 1 year)	Paix <i>et al.</i> <sup>72</sup>
SABR complications	0.108 (per year)	Paix <i>et al.</i> <sup>72</sup>
Lobectomy and wedge-resection complications	0.042 (per 6 months)	Paix <i>et al.</i> <sup>72</sup>

## Appendix 14 Diagnostic yield and accuracy of lung biopsy

Source	Procedure	Value	Studies (n)	References
Ghambir <i>et al.</i> <sup>22</sup> 2000	TTNAB	0.895/0.959	6	<ul style="list-style-type: none"> <li>Dewan N, Reeb S, Gupta N. PET-FDG imaging and transthoracic needle lung aspiration biopsy in evaluation of pulmonary lesions. <i>Chest</i> 1995;<b>108</b>:441–6</li> <li>Khoury N, Stitik F. Transthoracic needle aspiration biopsy of benign and malignant lung lesions. <i>AJR</i> 1985;<b>144</b>:281–8</li> <li>Stanley JH, Fish GD, Andriole JG, Gobien RP, Betsill WL, Laden SA, <i>et al.</i> Lung lesions: cytologic diagnosis by fine-needle biopsy. <i>Radiology</i> 1987;<b>162</b>:389–91</li> <li>Charig MJ, Stutley JE, Padley SP, Hansell DM. The value of negative needle biopsy in suspected operable lung cancer. <i>Clin Radiol</i> 1991;<b>44</b>:147–9</li> <li>Cavouras D, Prassopoulos P, Pantelidis N. Image analysis methods for solitary pulmonary nodule characterization by computed tomography. <i>Eur J Radiol</i> 1992;<b>14</b>:169–72</li> <li>Grode G, Faurischou P, Milman N. Percutaneous transthoracic fine-needle lung biopsy with 3 different needles. A retrospective study of results and complications in 224 patients. <i>Respiration</i> 1993;<b>60</b>:284–8</li> </ul>
Dietlein <i>et al.</i> <sup>24</sup> 2000	TNB	0.900/1.000	4	<ul style="list-style-type: none"> <li>Dewan N, Reeb S, Gupta N. PET-FDG imaging and transthoracic needle lung aspiration biopsy in evaluation of pulmonary lesions. <i>Chest</i> 1995;<b>108</b>:441–6</li> <li>Khoury N, Stitik F. Transthoracic needle aspiration biopsy of benign and malignant lung lesions. <i>AJR</i> 1985;<b>144</b>:281–8</li> <li>Klein JS, Zarka MA. Transthoracic needle biopsy: an overview. <i>J Thorac Imaging</i> 1997;<b>12</b>:232–49</li> <li>Larscheid RC, Thorpe PE, Scott WJ. Percutaneous transthoracic needle aspiration biopsy: a comprehensive review of its current role in the diagnosis and treatment of lung tumors. <i>Chest</i> 1998;<b>114</b>:704–9</li> </ul>
Keith <i>et al.</i> <sup>21</sup> 2002		0.95/0.88	2	<ul style="list-style-type: none"> <li>Institute of Clinical PET Solitary Nodule Task Force. Clinical application and economic implications of PET in the assessment of solitary pulmonary nodules: a retrospective study. Foothill Ranch, CA: ICP; 1994</li> </ul>
		0.90/0.96		<ul style="list-style-type: none"> <li>Gambhir SS, Shephard JE, Shah BD, Hart E, Hoh CK, Valk PE, <i>et al.</i> Analytical decision model for the cost-effective management of solitary pulmonary nodules. <i>J Clin Oncol</i> 1998;<b>16</b>:2113–25</li> </ul>
Comber <i>et al.</i> <sup>26</sup> 2003		0.90/0.96	1	<ul style="list-style-type: none"> <li>Gambhir SS, Shephard JE, Shah BD, Hart E, Hoh CK, Valk PE, <i>et al.</i> Analytical decision model for the cost-effective management of solitary pulmonary nodules. <i>J Clin Oncol</i> 1998;<b>16</b>:2113–25</li> </ul>
Gould <i>et al.</i> <sup>20</sup> 2003		0.963/0.980	9	<ul style="list-style-type: none"> <li>Dewan NA, Reeb SD, Gupta NC, Gobar LS, Scott WJ. PET-FDG imaging and transthoracic needle lung aspiration biopsy in evaluation of pulmonary lesions. A comparative risk-benefit analysis. <i>Chest</i> 1995;<b>108</b>:441–6</li> <li>Van Sonnenberg E, Casola G, Ho M, Neff CC, Varney RR, Wittich GR, <i>et al.</i> Difficult thoracic lesions: CT-guided biopsy experience in 150 cases. <i>Radiology</i> 1988;<b>167</b>:457–61</li> <li>Garcia Rio F, Diaz Lobato S, Pino JM, Atienza M, Viguer JM, Villasante C, <i>et al.</i> Value of CT-guided fine needle aspiration in solitary pulmonary nodules with negative fiberoptic bronchoscopy. <i>Acta Radiol</i> 1994;<b>35</b>:478–80</li> <li>Li H, Boiselle PM, Shepard JO, Trotman-Dickenson B, McLoud TC. Diagnostic accuracy and safety of CT-guided percutaneous needle aspiration biopsy of the lung: comparison of small and large pulmonary nodules. <i>AJR Am J Roentgenol</i> 1996;<b>167</b>:105–9</li> </ul>

Source	Procedure	Value	Studies (n)	References
				<ul style="list-style-type: none"> <li>• Santambrogio L, Nosotti M, Bellaviti N, Pavoni G, Radice F, Caputo V. CT-guided fine-needle aspiration cytology of solitary pulmonary nodules: a prospective, randomized study of immediate cytologic evaluation. <i>Chest</i> 1997;<b>112</b>:423–5</li> <li>• Westcott JL, Rao N, Colley DP. Transthoracic needle biopsy of small pulmonary nodules. <i>Radiology</i> 1997;<b>202</b>:97–103</li> <li>• Yankelevitz DF, Henschke CI, Koizumi JH, Altorki NK, Libby D. CT-guided transthoracic needle biopsy of small solitary pulmonary nodules. <i>Clin Imaging</i> 1997;<b>21</b>:107–10</li> <li>• Hayashi N, Sakai T, Kitagawa M, Kimoto T, Inagaki R, Ishii Y, <i>et al.</i> CT-guided biopsy of pulmonary nodules less than 3 cm: usefulness of the springoperated core biopsy needle and frozen-section pathologic diagnosis. <i>AJR Am J Roentgenol</i> 1998;<b>170</b>:329–31</li> <li>• Laurent F, Latrabe V, Vergier B, Montaudon M, Vernejoux JM, Dubrez J. CT-guided transthoracic needle biopsy of pulmonary nodules smaller than 20 mm: results with an automated 20-gauge coaxial cutting needle. <i>Clin Radiol</i> 2000;<b>55</b>:281–7</li> </ul>
Tsushima and Endo <sup>27</sup> 2004	CT-guided needle biopsy	0.769/0.936	1	<ul style="list-style-type: none"> <li>• Tsukada H, Satou T, Iwashima A, Souma T. Diagnostic accuracy of CT-guided automated needle biopsy of lung nodules. <i>AJR Am J Roentgenol</i> 2000;<b>175</b>:239–43</li> </ul>
Lejeune <i>et al.</i> <sup>28</sup> 2005	TNB	0.85/0.95	3	<ul style="list-style-type: none"> <li>• Greene R, Szyfel bein WM, Isler RJ, Stark P, Janstsch H. Supplementary tissue-core histology from fine-needle transthoracic aspiration biopsy. <i>AJR Am J Roentgenol</i> 1985;<b>144</b>:787–92</li> <li>• Charig MJ, Phillips AJ. CT-guided cutting needle biopsy of lung lesions-safety and efficacy of an out-patient service. <i>Clin Radiol</i> 2000;<b>55</b>:964–9</li> <li>• Dewan NA, Reeb SD, Gupta NC, Gobar LS, Scott WJ. PET-FDG imaging and transthoracic needle lung aspiration biopsy in evaluation of pulmonary lesions. A comparative risk-benefit analysis. <i>Chest</i> 1995;<b>108</b>:441–6</li> </ul>
Deppen <i>et al.</i> <sup>55</sup> 2014	CT-FNA	0.92/1.00, 77% yield	1	<ul style="list-style-type: none"> <li>• Rivera MP, Mehta AC. Initial diagnosis of lung cancer: ACCP evidence-based clinical practice guidelines (2nd edition). <i>Chest</i> 2007;<b>132</b>(Suppl. 1):131S–48S</li> </ul>

CT-FNA, computerised tomography-guided fine-needle aspiration; TTNAB, transthoracic needle aspiration biopsy.

Studies included in the systematic review of economic evaluations used a range of estimates for the sensitivity and specificity of biopsy, ranging from 0.769 to 0.963 for sensitivity, and 0.88 to 1.00 for specificity. There is little consistency in the diagnostic accuracy estimates in the studies included, and typically little detail is provided on the methods of searching for or of synthesising evidence. Three studies report using diagnostic accuracy estimates from other published sources: both Keith *et al.*<sup>21</sup> and Comber *et al.*<sup>26</sup> use values derived by Ghambir *et al.*,<sup>22</sup> with Keith *et al.*<sup>21</sup> also using values derived for a review undertaken by the Institute of Clinical PET Solitary Nodule Task Force. Deppen *et al.*<sup>55</sup> report using figures from an American College of Chest Physicians clinical guideline (however, the reported figures were not available in the cited reference).

The exception to this pattern is Gould *et al.*,<sup>20</sup> who report estimating diagnostic accuracy of CT-guided needle biopsy (and risk of complication) using studies identified by a MEDLINE search for English-language studies, published before January 2000, using terms for needle, biopsy, lung cancer and pulmonary nodules. Studies were included if they were limited to participants with pulmonary nodules (of < 4 cm) or reported these participants separately. Meta-analysis used the Moses–Shapiro–Littenberg method to construct a SROC curve, with base-case estimates derived as the mean sensitivity across included studies and specificity being the point on the SROC curve that matched this sensitivity estimate.

Because the most recent searches for evidence of diagnostic accuracy of biopsy reported in these studies date back to November 2001, we undertook targeted searches for studies reporting systematic reviews or meta-analysis of diagnostic accuracy of CT-guided biopsy for SPNs.

Finally, we reviewed the studies included in the pooled analysis of 11 studies of CT-guided biopsy diagnostic accuracy reported in the BTS guidelines.<sup>9</sup> In these guidelines, pooled sensitivity and specificity of 0.91 and 0.94 were reported, respectively.

These data were analysed using the R MADA library for meta-analysis of diagnostic accuracy studies, using the Reits function to conduct a bivariate meta-analysis using the Reitsma method (R package version 0.5.8, Meta-Analysis of Diagnostic Accuracy, URL: <https://CRAN.R-project.org/package=mada>, accessed 10 December 2020). The analysis yielded sensitivity and specificity estimates of 0.912 (95% CI 0.873 to 0.940) and 0.955 (95% CI 0.901 to 0.980), respectively. Figure 26 shows the SROC fit by this procedure, the point estimate and 95% confidence ellipse.

TABLE 55 Studies reported in the BTS guidelines, included in the meta-analysis

Study	Cases (n)	Useful sample (n)	True positive, n (malignant, n)	True negative, n (non-malignant, n)	Sensitivity/specificity
Baldwin <i>et al.</i> <sup>113</sup>	114	98	71 (74)	23 (24)	0.96/0.96
Gupta <i>et al.</i> <sup>114</sup>	176	143	104 (109)	34 (34)	0.95/1.00
Hayashi <i>et al.</i> <sup>115</sup>	52	50	34 (35)	15 (0)	0.97/1.00
Jin <i>et al.</i> <sup>116</sup>	71	61	35 (36)	25 (0)	0.97/1.00
Ohno <i>et al.</i> <sup>117</sup>	396	396	266 (296)	80 (100)	0.90/0.72
Romano <i>et al.</i> <sup>118</sup>	229	184	113 (128)	56 (56)	0.88/1.00
Santambrogio <i>et al.</i> <sup>119</sup>	220	207	130 (138)	68 (69)	0.94/0.99
Tsukada <i>et al.</i> <sup>120</sup>	138	138	70 (91)	44 (47)	0.77/0.94
Wagnetz <i>et al.</i> <sup>121</sup>	108	104	79 (88)	16 (16)	0.90/1.00
Westcott <i>et al.</i> <sup>122</sup>	64	64	40 (43)	21 (21)	0.93/1.00

**Note**

Hayashi, Santambrogio and Westcott were included in the meta-analysis by Gould *et al.*;<sup>20</sup> Tsukada was used by Tsushima and Endo.<sup>27</sup>

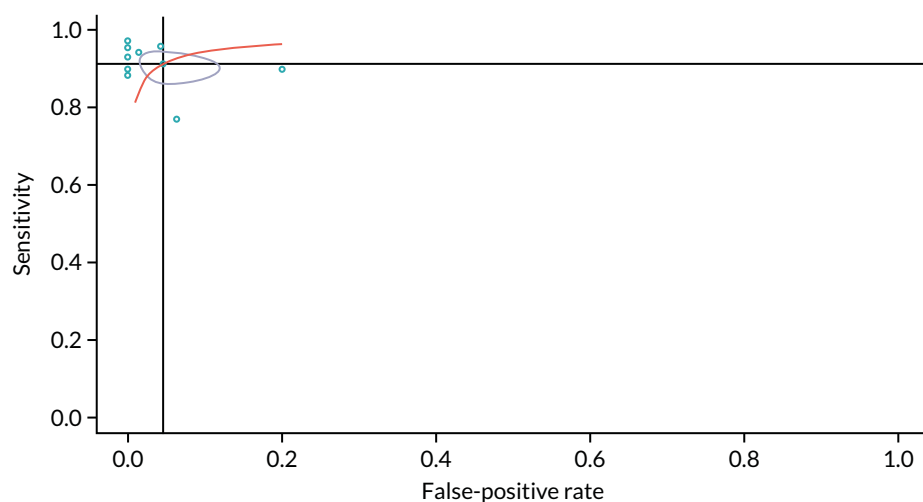


FIGURE 26 The SROC curve of biopsy diagnostic accuracy.



## Appendix 15 Biopsy and surgery complications

A targeted search strategy was used to identify additional evidence regarding the non-fatal and fatal complications of biopsy, surgery and radiotherapy.

TABLE 56 Inclusion/exclusion criteria and search terms

Criteria/search term	
<b>Inclusion criteria</b>	
Population	<ul style="list-style-type: none"> <li>Adults aged &gt; 19 years under investigation for a SPN</li> <li>'Solitary pulmonary nodule' as a major search-term heading</li> <li>SPN defined as nodules of &gt; 8 and &lt; 30 mm</li> </ul>
Intervention(s)	<ul style="list-style-type: none"> <li><sup>18</sup>F-FDG-PET/CT/CT/DCE-CT</li> <li>Biopsy: transthoracic needle biopsy or TNB or TTNB</li> <li>Surgery: thoracotomy/VAT or VATS or video-assisted thorascopy; lobectomy or wedge resection</li> </ul>
Setting	Secondary and tertiary care
Outcome measure	Procedure-related complications [incidence (n/N) or %]
Design	Primary research report: RCT/cohort/case control/cross sectional/prospective and retrospective
Other limits	<ul style="list-style-type: none"> <li>2000–18 then 2014–current</li> <li>English language</li> <li>Abstract and full text available</li> </ul>
<b>Exclusion criteria</b>	
Population	Pulmonary nodules of > 30 mm
Design	Single case studies
Other limits	<ul style="list-style-type: none"> <li>No full text available</li> <li>Not 2014 to current</li> </ul>
<b>Search terms</b>	
(solitary pulmonary nodule and (fluorodeoxyglucose positron emission tomography or surgery or biopsy or computed tomography or dynamic contrast enhanced) and (complications or problems or challenges or mortality or morbidity or death))	
VAT, video-assisted thorascopy; VATS, video-assisted thorascopic surgery.	

TABLE 57 Summary of biopsy complications reported in identified sources

Source	Procedure	Complications	Rate (n/N) <sup>a</sup>
Wang <i>et al.</i> <sup>123</sup> 2018	CT-PNB	Pneumothorax	0.175 (14/80)
		Haemorrhage	0.075 (6/80)
Xu <i>et al.</i> <sup>124</sup> 2018	CT-PTNB	Pneumothorax	0.161 (40/248)
		Haemorrhage	0.069 (17/248)
		Fatal adverse reactions	0.000 (0/248)
Liu <i>et al.</i> <sup>125</sup> 2017	MRI-guided percutaneous biopsy	Pneumothorax	0.123 (8/65)
		Haemoptysis	0.046 (3/65)

continued

TABLE 57 Summary of biopsy complications reported in identified sources (*continued*)

Source	Procedure	Complications	Rate (n/N) <sup>a</sup>
Liu <i>et al.</i> <sup>126</sup> 2015	MRI-guided percutaneous transthoracic needle biopsy	Pneumothorax	0.174 (12/69)
		Haemoptysis	0.072 (5/69)
Sa <i>et al.</i> <sup>127</sup> 2015	CT-guided PCNA biopsy with localisation for VATS	Pneumothorax	0.152 (5/33)
		Haemorrhage	0.091 (3/33)
Yang <i>et al.</i> <sup>128</sup> 2015	CT-guided TNB	Pneumothorax	0.177 (55/311)
		Haemorrhage	0.116 (36/311)
		Haemoptysis	0.035 (11/311)
Lee <i>et al.</i> <sup>129</sup> 2014	C-arm cone beam CT-PTNB	Pneumothorax	0.170 (196/1153)
		Haemoptysis	0.069 (80/1153)
Kaaki <i>et al.</i> <sup>130</sup> 2018	CT-guided needle biopsy	Pneumothorax	0.163 (8/49)
		Post-biopsy bleed	0.041 (2/49)
		Pneumothorax and bleed/haemoptysis	0.061 (3/49)

CT-PNB, computerised tomography-guided percutaneous needle biopsy; CT-PTNB, computerised tomography-guided percutaneous transthoracic needle biopsy; PCNA, percutaneous needle aspiration; VATS, video-assisted thoracoscopic surgery.  
a n/N = number of patients experiencing an event/number of patients in sample at risk of an event.

TABLE 58 Biopsy complication rates used in published economic evaluations (see *Chapter 6*)

Source	Procedure	Complication	Value	Sources (n) – pooling method clear?
Gambhir <i>et al.</i> <sup>22</sup> 1998	TNB	Pneumothorax with chest tube	0.240	6 – no
Dietlein <i>et al.</i> <sup>24</sup> 2000	TNB	Pneumothorax with chest tube	0.240	3 – no <sup>a</sup>
Gould <i>et al.</i> <sup>20</sup> 2003	TNB	Major pneumothorax	0.050	9 – yes (median value from studies)
		Minor pneumothorax	0.240	
Tsushima and Endo <sup>27</sup> 2004	CT-guided needle biopsy	Pneumothorax with chest tube	0.029	1 – NA
Lejeune <i>et al.</i> <sup>28</sup> 2005	TNB	Pneumothorax <sup>b</sup>	0.200	4 – no
Deppen <i>et al.</i> <sup>55</sup> 2014	CT-FNA	Pneumothorax needing observation	0.150	1 – NA
		Pneumothorax with chest tube	0.066	
		Haemorrhage	0.010	

CT-FNA, computerised tomography-guided fine-needle aspiration; NA, not applicable.  
a The authors include the references of three studies but appear to use values from one (Gambhir *et al.*<sup>22</sup>).  
b The published article does not indicate any distinction between all pneumothorax and those requiring chest tube.

TABLE 59 Summary of lobectomy and wedge-resection procedure-related complications reported in identified sources

Source	Procedure	Complications	Rate (n/N) <sup>a</sup>
Lu <i>et al.</i> <sup>131</sup> 2018	Concomitant VATS with modified epicardial radiofrequency ablation procedure for NVAf and SPN resection	Lobectomy: pneumothorax	? (?/13)
		Wedge resection: pneumothorax	? (?/3)
Kaaki <i>et al.</i> <sup>130</sup> 2018	Intraoperative wedge biopsy followed by lobectomy	Significant haemorrhage	0.077 (3/39)
Yang <i>et al.</i> <sup>132</sup> 2017	Tubeless uniportal thoracoscopic wedge resection	Tubeless	
		Pneumothorax	0.400 (12/30)
		Post-operative chest drainage (days)	0.0
		Chest tube	
		Pneumothorax	0.133 (4/30)
		Post-operative chest drainage (days)	1.7
Li <i>et al.</i> <sup>133</sup> 2017	Tubeless VATS	Pneumothorax	Not reported
Qi <i>et al.</i> <sup>134</sup> 2017	VATS combined with CT-guided dual-barbed hook wire localisation	Pneumothorax	0.032 (1/31)
		Parenchymal haemorrhage	0.161 (5/31)
Muller <i>et al.</i> <sup>135</sup> 2016	Handheld SPECT-navigated VATS of CT-guided radioactively marked pulmonary lesions	Pneumothorax	0.500 (5/10)
		Haemorrhage	0.100 (2/10)
Ghaly <sup>136</sup> 2016	Anatomic segmentectomy with thoracotomy/VATS	VATS	
		Pneumothorax	Not reported
		Haemorrhage	Not reported
		Air leak	0.077 (7/91)
		Thoracotomy	
		Pneumothorax	Not reported
		Haemorrhage	Not reported
Air leak	0.147 (15/102)		
Gill <i>et al.</i> <sup>137</sup> 2015	Image-guided VATS	Prolonged air leak	0.043 (1/23)
		Pneumonia	0.043 (1/23)
		Ileus	0.043 (1/23)
		Deaths	0.000 (0/23)
Ren <i>et al.</i> <sup>138</sup> 2014	VATS segmentectomy and VATS lobectomy	VATS segmentectomy	
		Pneumothorax	Not reported
		Haemorrhage	Not reported
		Atrial fibrillation	0.048 (1/21)
		Respiratory failure	0.048 (1/21)

continued

TABLE 59 Summary of lobectomy and wedge-resection procedure-related complications reported in identified sources (continued)

Source	Procedure	Complications	Rate (n/N) <sup>a</sup>
		VATS lobectomy	
		Pneumothorax	Not reported
		Haemorrhage	Not reported
		Atrial fibrillation	0.033 (2/61)
		Respiratory failure	0.016 (1/61)
		Pneumonia	0.016 (1/61)
		Vocal cord paralysis	0.016 (1/61)
Cho <i>et al.</i> <sup>139</sup> 2014	VATS	Pneumothorax	0.006 (2/330)
		Cyclothorax	0.003 (1/330)
		Pleural effusion	0.012 (4/330)
		Prolonged air leak	0.045 (15/330)

NVAF, non-valvular atrial fibrillation; SPECT, single-photon emission computerised tomography; VATS, video-assisted thoracoscopic surgery.

a n/N = number of patients experiencing an event/number of patients in sample at risk of an event.

TABLE 60 Surgical complication and mortality rates in published economic evaluations

Source	Procedure	Complication	Value	Sources (n) – pooling method clear?
Dietlin <i>et al.</i> <sup>24</sup> 2000	Surgical resection	Non-fatal complication	–	1 – NA
		Mortality	0.029	3 – no
Gould <i>et al.</i> <sup>20</sup> 2003	VATS	Non-fatal complication	0.065	4 – yes (weighted mean)
		Mortality	0.005	? – no
	Lobectomy	Non-fatal complication	0.084	4
		Mortality	0.042	4
Lejeune <i>et al.</i> <sup>28</sup> 2005	VATS	Non-fatal complication	0.2	2 – no
Deppen <i>et al.</i> <sup>55</sup> 2014	Wedge	Non-fatal complication	0.046	1 – NA
		Mortality	0.012	1 – NA
	Lobectomy	Non-fatal complication	0.064	1 – NA
		Mortality	0.022	1 – NA

VATS, video-assisted thoracoscopic surgery.

TABLE 61 Search strategy for radiotherapy complications

#	Search	Results (n)
1	(Radiosurgery or radiation therapy or SBRT or SABR).mp. [mp=ti, ab, ot, nm, hw, fx, kf, ox, px, rx, ui, sy, tn, dm, mf, dv, kw, dq]	225,103
2	(Lung neoplasm\$ or lung carcinoma\$ or (pulmonary adj2 nodule\$) or lung nodule\$ or non-small-cell lung or lung carcinoma\$).mp. [mp=ti, ab, ot, nm, hw, fx, kf, ox, px, rx, ui, sy, tn, dm, mf, dv, kw, dq]	385,268
3	(complication\$ or adverse effect\$ or adverse event\$ or toxicity or pneumonitis).mp. [mp=ti, ab, ot, nm, hw, fx, kf, ox, px, rx, ui, sy, tn, dm, mf, dv, kw, dq]	8,048,417
4	(meta-analys* or pooled analys*).mp. [mp=ti, ab, ot, nm, hw, fx, kf, ox, px, rx, ui, sy, tn, dm, mf, dv, kw, dq]	444,227
5	1 and 2 and 3 and 4	200
6	limit 5 to english language	193
7	limit 6 to human	183
8	limit 7 to humans	183
9	limit 8 to yr="2010 -Current"	135



## Appendix 16 Data used in sensitivity analyses

TABLE 62 Imaging test accuracy for scenario analysis

Description	Average	Probability distribution
<b><i>Indeterminate treated as positive</i></b>		
DCE-CT sensitivity	0.9785	Dirichlet (4, 182, 27, 96)
DCE-CT specificity	0.2195	
<sup>18</sup> F-FDG-PET/CT sensitivity	0.9788	Dirichlet (4, 185, 27, 96)
<sup>18</sup> F-FDG-PET/CT specificity	0.2195	
<b><i>Accuracy-based pre-defined thresholds</i></b>		
DCE-CT sensitivity (alt)	0.9524	Dirichlet (9, 180, 35, 87)
DCE-CT specificity (alt)	0.2868	
<sup>18</sup> F-FDG-PET/CT sensitivity (alt)	0.7513	Dirichlet (47, 142, 97, 23)
<sup>18</sup> F-FDG-PET/CT specificity (alt)	0.8083	



## Appendix 17 The IPCARD-SPN questionnaire substudy: additional tables

TABLE 63 Baseline characteristics for the SPUtNik study subsample with valid IPCARD-SPN questionnaire data and valid lung cancer outcome data

Variable (unit)	N = 281
<b>Sex, n (%)</b>	
Female	128 (46)
Male	150 (53)
Missing	3 (1)
<b>Age (years)</b>	
Mean	68.22
Median	69
LQ to UQ	62.25 to 89
Minimum, maximum	35, 89
Missing, n (%)	3 (1)
<b>Smoking status (SPUtNik medical history), n (%)</b>	
Non-smoker	51 (18)
Ex-smoker	134 (48)
Current smoker	6 (2)
Missing	90 (32)
<b>Years smoked (IPCARD-SPN questionnaire)</b>	
Mean	30.13
Median	36
LQ to UQ	10 to 46
Minimum, maximum	0, 71
Missing, n (%)	36 (13)
<b>Medical history of cardiovascular disease, n (%)</b>	
Any cardiovascular disease	63 (22)
No cardiovascular disease	204 (73)
Missing	14 (5)
<b>Medical history of respiratory disease, n (%)</b>	
Any respiratory disease	113 (40)
No respiratory disease	160 (57)
Missing	8 (3)
COPD	73 (64)
Missing	168 (60)
	continued

TABLE 63 Baseline characteristics for the SPUTNIK study subsample with valid IPCARD-SPN questionnaire data and valid lung cancer outcome data (continued)

Variable (unit)	N = 281
<b>Previous exposures, n (%)</b>	
Any previous exposure (asbestos, coal or silica)	49 (17)
No previous exposure	185 (66)
Missing	47 (17)
<b>Prior malignancy, n (%)</b>	
Yes	40 (14)
No	233 (83)
Missing	8 (3)
<b>Nodule size (mm)</b>	
Mean	15.61
Median	14.50
LQ to UQ	12 to 19
Minimum, maximum	1.5, 30
Missing, n (%)	5 (2)
LQ, lower quartile; UQ, upper quartile.	

TABLE 64 Predicted probabilities for the diagnosis of lung cancer within the first 12 months of surveillance (model 1)

Symptoms	Never		Within the previous 3 months		> 3 months ago		Never/first experienced > 3 months ago		First experienced within the previous 3 months	
	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
Sweats not caused by the menopause	0.64	0.49	0.23	0.14	0.55	0.40	-	-	-	-
Unexpected tiredness	-	-	-	-	-	-	0.64	0.49	0.87	0.80



EME  
HSDR  
**HTA**  
PGfAR  
PHR

Part of the NIHR Journals Library  
[www.journalslibrary.nihr.ac.uk](http://www.journalslibrary.nihr.ac.uk)

*This report presents independent research funded by the National Institute for Health Research (NIHR).  
The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the  
Department of Health and Social Care*

***Published by the NIHR Journals Library***