National Institute for Health and Care Excellence

Final

Lung cancer update

[B] Evidence reviews for the clinical and cost-effectiveness of routine MRI or CT of the brain in the management of people with lung cancer prior to therapy with curative intent

NICE guideline NG122

Evidence reviews

March 2019

Final

These evidence reviews were developed by the NICE Guideline Updates Team

Disclaimer

The recommendations in this guideline represent the view of NICE, arrived at after careful consideration of the evidence available. When exercising their judgement, professionals are expected to take this guideline fully into account, alongside the individual needs, preferences and values of their patients or service users. The recommendations in this guideline are not mandatory and the guideline does not override the responsibility of healthcare professionals to make decisions appropriate to the circumstances of the individual patient, in consultation with the patient and/or their carer or guardian.

Local commissioners and/or providers have a responsibility to enable the guideline to be applied when individual health professionals and their patients or service users wish to use it. They should do so in the context of local and national priorities for funding and developing services, and in light of their duties to have due regard to the need to eliminate unlawful discrimination, to advance equality of opportunity and to reduce health inequalities. Nothing in this guideline should be interpreted in a way that would be inconsistent with compliance with those duties.

NICE guidelines cover health and care in England. Decisions on how they apply in other UK countries are made by ministers in the Welsh Government, Scottish Government, and Northern Ireland Executive. All NICE guidance is subject to regular review and may be updated or withdrawn.

Copyright

ISBN: 978-1-4731-3307-5

Contents

Evidence reviews for the clinical and cost-effectiveness of routine MRI or CT of the brain in the management of people with lung cancer prior to radical therapy with curative intent

Review questions

RQ1.3: What is the clinical and cost-effectiveness of routine MRI or CT of the brain in the management of people with lung cancer prior to radical therapy with curative intent?

Introduction

This area was identified for review because observational data calculating the prevalence of brain metastases in people with various stages of NSCLC selected for treatment with curative intent has been published since the last guideline (O'Dowd 2014[1]). This data enabled the effectiveness and cost-effectiveness of various imaging strategies to be calculated. The 2011 NICE lung cancer guideline recommended that MRI or CT scan should be considered before treatment with curative intent, especially for (patients otherwise thought to have) stage III NSCLC. MRI brain may be more accurate at detecting brain metastases compared to CT brain. However, there is reduced availability and increased cost for MRI compared to CT. The prevalence of brain metastases is likely to by population subgroup.

Table 1: PICO table

Population	People with lung cancer stage I to stage IIIA considered for radical treatment
Intervention	MRI brain or CT brain
Comparator	No imaging. Brain metastases identified during follow up period
Outcomes	• Diagnostic sensitivity and specificity (likelihood ratios) • Staging sensitivity and specificity • Safety of each procedure/ adverse events • Patient acceptability • Anxiety and psychological outcomes • Change in treatment plan • Change in staging

Methods and process

This evidence review was developed using the methods and process described in Developing NICE guidelines: the manual (2014). Methods specific to this review question are described in the review protocol in appendix A, and the methods section in appendix B. In particular, the minimally important differences (MIDs) used in this review are summarised in appendix B.

[1] O'Dowd et al (2014) Brain metastases following radical surgical treatment of non-small cell lung cancer: is preoperative brain imaging important? Lung Cancer. 2014 Nov;86(2):185-9

There was a deviation from the protocol: for the diagnostic test accuracy outcomes (sensitivity and specificity), the population of interest was increased from NSCLC stages I-IIIA to include all stages (NSCLC stages IIIB and IV too). This was to obtain more accuracy data for imaging to detect brain metastases. Studies that only have participants with NSCLC stages I-IIIA are few and have relatively low numbers of participants. Where we used studies that included participants with stages IIIB and IV, we downgraded for indirectness.

A consultant neuroradiologist was co-opted onto the committee to provide advice and expertise for this research question. Declarations of interest were recorded according to NICE's 2018 conflicts of interest policy.

Clinical evidence

Included studies

This review was conducted as part of a larger update of the NICE Lung cancer: diagnosis and management guideline (CG121). A systematic literature search for randomised controlled trials (RCTs), systematic reviews of RCTs and observational studies including cohort trials with a no date limit yielded 4,216 references.

Papers returned by the literature search were screened on title and abstract, with 24 full-text papers ordered as potentially relevant references.

Nine papers representing 9 unique studies were included after full text screening:

Table of included studies

Study	Study type	Intervention (s)	Number of participants	Follow-up period	Study location
Earnest 1999	Prospective cohort study	MRI brain	138	12 months	USA
Hochstenbag 2003	Prospective cohort study	MRI brain	51	At least 6 months	The Netherlands
Kim 2005	Prospective cohort study	MRI brain	69	No follow-up period	South Korea
Kormas 1992	Prospective cohort study	CT brain	158	12 months	UK
Lee 2009	Prospective cohort study	MRI brain	442	6 months	South Korea
Ferrigno 1994	Retrospective study	CT brain	184	12 months	Italy
de Cos Escuin 2007	Retrospective study	MRI brain or CT brain	170	Follow-up was a cut-off between 3 to 17 months	Spain

Yohena 2004	Retrospective study	MRI brain	127	No follow-up period	Japan
Yokoi 1999	Retrospective study	MRI brain or CT brain	332	12 months	Japan

For the search strategy, please see appendix C. For the clinical evidence study selection flowchart, see appendix D. For the full evidence tables and full GRADE profiles for included studies, please see appendix E and appendix F.

Excluded studies

Details of the studies excluded at full-text review are given in appendix H along with a reason for their exclusion.

Summary of clinical studies included in the evidence review

Outcomes and sample sizes

See full evidence tables and Grade profiles Appendix E and Appendix F.

Quality assessment of clinical studies included in the evidence review

See appendix F for full GRADE tables.

Economic evidence

Standard health economic filters were applied to the clinical search for this question, and a total of 401 citations was returned. Details of the literature search are provided in Appendix C. Following review of titles and abstracts, 2 full-text studies were retrieved for detailed consideration, of which none were included in our review.

Summary of original economic model

The de-novo cost-utility analysis developed for this guideline (see Appendix I for full details) included three strategies; no imaging (i.e. proceed straight to treatment with curative intent), imaging with CT brain, followed by MRI brain if positive and imaging with MRI brain. Patients in the model were divided into three categories; negative, positive with 1-3 brain metastases and positive with 4+ metastases. These were decided upon as the most clinically relevant patient groups. The model examined patients with NSCLC stage I, stage II and stage IIIA separately. Patients found to be negative exited the model because the tests were assumed (based on the evidence identified and the committee's experience) to have a specificity of 100%. CT and MRI were also assumed to have a sensitivity of 100% for detecting 4+ metastases in the model's base case. After imaging or no imaging, patients could therefore be true positive with 1-3 brain metastases, true positive with 4+, false negative with 1-3 or undetected with 4+. This final group only existed in the no imaging strategy in the base case. Following detection of brain metastases, radical treatments shifted from more to less invasive techniques and radical treatments were assumed to be used less frequently. Patients also received appropriate treatment for their brain

metastases. After initial imaging and treatment, patients entered the long term part of the model where their overall and progression-free survival was modelled using data from relevant RCTs and cohort studies. Patients received indicated treatments upon progression and death.

The model found that imaging was not cost-effective in stage I NSCLC, that CT followed by MRI if positive could be cost-effective in stage II disease and MRI was the dominant strategy (the cheapest and most effective) in stage IIIA disease. These results were robust to plausible sensitivity and scenario analyses. The most important parameters in the model were the prevalence of brain metastases, the proportion of positives who had 4+ metastases and the extent to which the treatment plan was assumed to change following initial imaging.

Evidence statement

MRI brain

Diagnostic accuracy data: meta-analysis

Very low-quality evidence from 4 observational studies on 624 people with stage I to stage IV lung cancer considered for radical treatment found that for MRI brain the sensitivity was 94.1% (68.6 – 99.9) and the specificity was 99.9% (91.0 – 100.0).

Diagnostic accuracy data: Yokoi 1999

Very low-quality evidence from 1 observational study on 177 people with stage I to stage IV lung cancer considered for radical treatment found that for MRI brain the sensitivity was 50% (26.1 – 73.9) and the specificity was 99.7% (97.2 – 100). This data was excluded from the meta-analysis above due to clinical implausibility; the sensitivity was too low. This was a post hoc decision by the guideline committee.

Effectiveness data (change in treatment plan: initially operable people who had metastases detected by imaging)

Very low-quality evidence from 4 prospective cohort studies and 1 retrospective cohort study reporting data on 558 people with stage I to stage IIIA lung cancer considered for radical treatment found that the percentage who were found to have brain metastases using MRI brain ranged from 1.5% (CI 0.19% - 5.57%) to 21.4% (8.3% - 31%).

CT brain with contrast

Diagnostic accuracy data: meta-analysis

Very low-quality evidence from 3 observational studies on 418 people with stage I to stage IV lung cancer considered for radical treatment found that for CT brain the sensitivity was 74.6% (11.5 – 99.7) and the specificity was 99.7% (85.2 – 100.0).

Diagnostic accuracy data: Yokoi 1999

Very low-quality evidence from 1 observational study on 177 people with stage I to stage IV lung cancer considered for radical treatment found that for CT brain the sensitivity was 12.5% (2.9 – 40.2) and the specificity was 99.7% (96.8 – 100). This

data was excluded from the meta-analysis above due to clinical implausibility; the sensitivity was too low. This was a post hoc decision by the guideline committee.

Effectiveness data (change in treatment plan: initially operable people who had metastases detected by imaging)

Very low-quality evidence from 1 prospective cohort study reporting data on 152 people with stage I to stage IIIA lung cancer considered for radical treatment found that the percentage who were found to have brain metastases using CT brain was 6.29% (2.92 – 11.6).

Health economics evidence statement

Evidence from one directly applicable health economic model with minor limitations developed for this guideline found that brain imaging was not cost-effective in patients with stage I NSCLC otherwise being considered for treatment with curative intent. The model found that a strategy of CT followed by MRI if positive was the most cost-effective for stage II disease at a threshold of £30,000/QALY and might have been cost-effective at a threshold of £20,000/QALY. MRI alone was the most cost-effective strategy in stage III disease.

The committee's discussion of the evidence

Interpreting the evidence

The outcomes that matter most

The committee agreed that the outcome that matters most is not causing harm by offering treatment options with curative intent, particularly surgical options, in patients who have brain metastases. Radical treatment options for lung cancer are associated with risks, side effects, high healthcare resource use and are not expected to alter the prognosis of many people with brain metastases. Another important outcome is the potential benefit of being able to offer alternative treatments to patients who have brain metastases. Early identification and appropriate management may slow disease progression and increase overall survival.

The quality of the evidence

The quality of the evidence included in the clinical review was very low. The committee noted that there is no agreed gold standard for assessing the presence of brain metastases and therefore the data on sensitivity in the included studies is particularly unreliable. The original health economic model developed for this review question included a large amount of evidence of varying quality, including a large number of assumptions and extrapolations from indirect data but overall the committee considered it a robust analysis for decision making. This was because its conclusions for each disease stage were not sensitive to plausible variations in any of the input parameters. In the meta-analyses of sensitivity and specificity data for MRI and CT brain, we excluded Yokoi 1999 from the analysis because the sensitivity data in this study are implausible compared to the sensitivity of modern MRI and CT brain imaging.

Benefits and harms

Imaging of the brain for those being considered for surgery or radical radiotherapy should prevent the use of radical treatment options in some patients for whom it is not expected to be beneficial. In addition, patients found to have brain metastases could be considered for other treatments such as stereotactic radiosurgery or brain surgery, which would be expected to improve their prognosis although each treatment would carry its own risks and side effects. The committee agreed that some patients feel anxiety on undergoing MRI but agreed that the scan was a safe and highly accurate way to detect brain metastases.

Cost-effectiveness and resource use

The recommendations for this area were based on the health economic model developed for this update (see Appendix I). The economic model examined three strategies; no imaging, CT (followed by MRI if the CT was positive) and MRI alone in patients with stage I, stage II and stage IIIA NSCLC being considered for treatment with curative intent separately. Early identification of brain metastases within the model led to an increase in Quality Adjusted Life Years (QALYs) because earlier management of brain metastases led to slower rates of progression and higher overall survival. There were costs associated with the initial imaging and subsequent treatment of brain metastases but also some savings from patients receiving less radical treatment, particularly surgery. Broadly, there were two types of patients within the economic model, those with 1-3 brain metastases, many of whom would receive radical treatment for their primary tumour as well as their metastases, and those with 4+ metastases who were modelled to no longer receive radical treatment but to move to systemic therapy. The committee were aware there would be some exceptions to these groupings in practice but felt the split was clinically meaningful and that it was a distinction that had often been made in the evidence base. Because of the associated cost savings, the proportion of positive patients who have 4+ brain metastases was an important but uncertain parameter in the economic model. The committee noted this uncertainty in its interpretation of the evidence for different stages of NSCLC.

To calculate test outcomes in the model, a diagnostic test accuracy meta-analysis was undertaken. This found that the sensitivity of CT and MRI were 74% and 94% respectively and that both modalities had a specificity of ~100%. The committee thought this was reasonable, particularly in relation to MRI so there were no patients with a false diagnosis of brain metastases included in the economic model. While the prevalence would likely be affected by the mixed population in some of the studies, the committee did not think the sensitivity of the tests would be and understood that these values would be thoroughly tested in scenario analysis in the model.

The evidence on the prevalence of brain metastases within the model came from a retrospective cohort analysis that had extrapolated data on patients treated with curative intent who had subsequently developed brain metastases. The authors of this paper used tumour doubling times to calculate how many patients would have had detectable brain metastases at the time of their radical treatment. The committee understood the limitations of this kind of analysis but also considered it to be the best available source of evidence that was relevant to the decision problem. The paper reported the estimated prevalence for stages I, II and IIIA separately.

The base case ICERs for CT-MRI versus No Imaging and MRI versus CT-MRI in stage I patients were greater than £30,000/QALY gained. There were no sensitivity analyses that moved these values close to £20,000/QALY gained. This was primarily because of the low prevalence of brain metastases in Stage I patients. The committee also noted that for every 100 MRI scans performed, only 3 patients would be found positive for brain metastases. They therefore decided that it was highly unlikely that imaging in Stage I represented a cost-effective use of NHS resources.

The base case ICERs for CT-MRI versus No Imaging and MRI versus CT-MRI in stage II patients were £21,000/QALY and £48,000/QALY respectively. There were no plausible sensitivity analyses that made MRI cost-effective compared to CT-MRI. The primary reasons for these findings are that CT was assumed to have very good sensitivity for identifying patients who have 4+ brain metastases and these patients are the most important in the cost-effectiveness calculations within the model because they are no longer likely to receive radical treatment, leading to significant cost savings. The committee noted that only a small number of people with 1-3 brain metastases would be missed on initial CT that might have been detected had MRI been the first test. They therefore decided to recommend a strategy of CT, followed by MRI if positive in the Stage II NSCLC population.

For stage IIIA patients, MRI was the dominant strategy (it was both cheaper and more effective) and remained either dominant or the most cost-effective strategy in all plausible sensitivity analyses. This is because all stage IIIA patients found to be positive for brain metastases are highly unlikely to receive radical treatment, leading to significant cost savings in the model. These savings, coupled with the relatively high prevalence of brain metastases and the clinical benefits of early diagnosis mean that the most sensitive test, MRI, is the most cost-effective.

The committee noted a number of limitations in the economic model relating to its data inputs and assumptions but also noted the findings were robust to all plausible sensitivity analyses and were therefore confident that it was reliable as the basis for decision making for this review question.

Other factors the committee took into account

The committee was aware that there are pressures on imaging services, particularly MRI scanners and that some patients prefer not to receive MRI scans but agreed that these considerations should not affect the recommendations. Some of the evidence that underpinned the health economic model was of low quality or based on committee assumption. In particular, they considered that due to the non-contemporary nature of the studies, the sensitivity of CT and MRI are likely to be underestimated with the use of thin collimation and volumetric imaging having improved the accuracy of both modalities in recent years. The committee was satisfied that these concerns had been addressed by an extensive range of sensitivity analyses. The main evidence for the prevalence of brain metastases came from a paper where the population of interest had not received contrast enhanced PET-CT as part of their staging. The committee acknowledged that in centres where contrast enhanced PET-CT is routine, the prevalence of brain metastases in the population of interest might be lower. While the specificity of MRI was thought to be 100% as regards brain metastases from lung cancer, the committee noted that several differential diagnoses such as infection and primary brain tumour might be detected by the scan. They considered this an ancillary benefit of imaging.

For recommendations 1.3.23, 1.3.24, and 1.3.25 the committee agreed that 'clinical stage' should be written, rather than 'stage'. This is to ensure that healthcare professionals understand that the brain imaging should be performed before surgery. After surgery, a pathologist is able to confirm the stage. Other recommendations do not require this clarification because it is normally obvious to clinicians whether a stage is clinical or not.

For recommendation 1.3.25, the committee agreed that the stage should be clinical stage III, which includes IIIA and IIIB. This is because some people with stage IIIB disease will receive radical radiotherapy and it is highly likely that MRI brain would be just as cost-effective in these patients as in patients with stage IIIA NSCLC.

Appendix A – Review protocols

Review protocol for the clinical and cost-effectiveness of routine MRI or CT of the brain in the management of people with lung cancer prior to radical therapy with curative intent?

Field (based on PRISMA-P)	Content
Review question	What is the clinical and cost-effectiveness of routine MRI or CT of the brain in the management of people with lung cancer prior to radical therapy with curative intent?
Type of review question	Intervention or diagnostic accuracy
Objective of the review	To assess whether the recommendation to 'consider MRI or CT of the head in patients selected for treatment with curative intent' requires updating. This area was identified during the scoping phase of the update. Variation in practice has also been identified.
Eligibility criteria – population	Patients with stage I to stage IIIA considered for radical treatment. For

	diagnostic test accuracy outcomes (sensitivity and specificity), patients with all stages
Eligibility criteria – interventions	MRI brain or CT brain
Eligibility criteria – comparator	Brain metastases identified during follow up period
Outcomes and prioritisation	• Diagnostic sensitivity and specificity • Staging sensitivity and specificity • Safety of each procedure/ adverse events • Patient acceptability • Anxiety and psychological outcomes • Change in treatment plan • Change in staging
Eligibility criteria – study design	• RCTs • Systematic reviews of RCTs • Observational studies including cohort trials

Other inclusion exclusion criteria	• Non- English-language papers • Unpublished evidence/ conference proceedings
Proposed sensitivity/sub-group analysis, or meta-regression	Stage I vs Stage II vs Stage IIIA
Selection process – duplicate screening/selection/analysis	10% of the abstracts were reviewed by two reviewers, with any disagreements resolved by discussion or, if necessary, a third independent reviewer. If meaningful disagreements were found between the different reviewers, a further 10% of the abstracts were reviewed by two reviewers, with this process continued until agreement is achieved between the two reviewers. From this point, the remaining abstracts will be screened by a single reviewer. This review made use of the priority screening functionality with the EPPI-reviewer systematic reviewing software. See Appendix B for more details.
Data management (software)	See Methods Appendix B
Information sources – databases and dates	See Appendix C Main Searches:

	• Cochrane Database of Systematic Reviews – CDSR • Cochrane Central Register of Controlled Trials – CENTRAL • Database of Abstracts of Reviews of Effects – DARE • Health Technology Assessment Database – HTA • EMBASE (Ovid) • MEDLINE (Ovid) • MEDLINE In-Process (Ovid) Citation searching will be carried out in addition on analyst/committee selected papers. The search will not be date limited because this is a new review question. Economics: • NHS Economic Evaluation Database – NHS EED • Health Economic Evaluations Database – HEED • EconLit (Ovid) • Embase (Ovid) • MEDLINE (Ovid) • MEDLINE In-Process (Ovid)

	The search will not be date limited because this is a new review question.
Identify if an update	This is not an update, it is a new review question.
Author contacts	Guideline update
Highlight if amendment to previous protocol	For details please see section 4.5 of Developing NICE guidelines: the manual
Search strategy – for one database	For details please see appendix C
Data collection process – forms/duplicate	A standardised evidence table format will be used, and published as appendix F (clinical evidence tables) or I (economic evidence tables).
Data items – define all variables to be collected	For details please see evidence tables in appendix F (clinical evidence tables) or I (economic evidence tables).
Methods for assessing bias at outcome/study level	See Appendix B
Criteria for quantitative synthesis	See Appendix B

Methods for quantitative analysis – combining studies and exploring (in)consistency	See Appendix B
Meta-bias assessment – publication bias, selective reporting bias	See Appendix B
Confidence in cumulative evidence	See Appendix B
Rationale/context – what is known	For details please see the introduction to the evidence review in the main file.
Describe contributions of authors and guarantor	A multidisciplinary committee developed the evidence review. The committee was convened by the NICE Guideline Updates Team and chaired by Gary McVeigh in line with section 3 of Developing NICE guidelines: the manual. Staff from the NICE Guideline Updates Team undertook systematic literature searches, appraised the evidence, conducted meta-analysis and cost-effectiveness analysis where appropriate, and drafted the evidence review in collaboration with the committee. For details please see Developing NICE guidelines: the manual.

Sources of funding/support	The NICE Guideline Updates Team is an internal team within NICE.
Name of sponsor	The NICE Guideline Updates Team is an internal team within NICE.
Roles of sponsor	The NICE Guideline Updates Team is an internal team within NICE.
PROSPERO registration number	N/A

Appendix B – Methods

Priority screening

The reviews undertaken for this guideline all made use of the priority screening functionality with the EPPI-reviewer systematic reviewing software. This uses a machine learning algorithm (specifically, an SGD classifier) to take information on features (1, 2 and 3 word blocks) in the titles and abstract of papers marked as being 'includes' or 'excludes' during the title and abstract screening process, and re-orders the remaining records from most likely to least likely to be an include, based on that algorithm. This re-ordering of the remaining records occurs every time 25 additional records have been screened.

Research is currently ongoing as to what are the appropriate thresholds where reviewing of abstract can be stopped, assuming a defined threshold for the proportion of relevant papers it is acceptable to miss on primary screening. As a conservative approach until that research has been completed, the following rules were adopted during the production of this guideline:

- In every review, at least 50% of the identified abstract (or 1,000 records, if that is a greater number) were always screened.
- After this point, screening was only terminated when the threshold was reached for a number of abstracts being screened without a single new include being identified. This threshold was set according to the expected proportion of includes in the review (with reviews with a lower proportion of includes needing a higher number of papers without an identified study to justify termination), and was always a minimum of 250.
- A random 10% sample of the studies remaining in the database when the threshold were additionally screened, to check if a substantial number of relevant studies were not being correctly classified by the algorithm, with the full database being screened if concerns were identified.

As an additional check to ensure this approach did not miss relevant studies, the included studies lists of included systematic reviews were searched to identify any papers not identified through the primary search.

Evidence synthesis and meta-analyses

Where possible, meta-analyses were conducted to combine the results of studies for each outcome. For mean differences, where change from baseline data were reported in the studies and were accompanied by a measure of spread (for example standard deviation), these were extracted and used in the meta-analysis. Where measures of spread for change from baseline values were not reported, the corresponding values at study end were used and were combined with change from baseline values to produce summary estimates of effect. All studies were assessed to ensure that baseline values were balanced across the treatment/comparison groups; if there were significant differences in important confounding variables at baseline these studies were not included in any meta-analysis and were reported separately.

When averages were given as medians, we presented them as they were.

Evidence of effectiveness of interventions

Quality assessment

Individual RCTs and quasi-randomised controlled trials were quality assessed using the Cochrane Risk of Bias Tool. Cohort studies were quality assessed using the CASP cohort study checklist. Each individual study was classified into one of the following three groups:

- Low risk of bias – The true effect size for the study is likely to be close to the estimated effect size.
- Moderate risk of bias – There is a possibility the true effect size for the study is substantially different to the estimated effect size.
- High risk of bias – It is likely the true effect size for the study is substantially different to the estimated effect size.

Each individual study was also classified into one of three groups for directness, based on if there were concerns about the population, intervention, comparator and/or outcomes in the study and how directly these variables could address the specified review question. Studies were rated as follows:

- Direct – No important deviations from the protocol in population, intervention, comparator and/or outcomes.
- Partially indirect – Important deviations from the protocol in one of the population, intervention, comparator and/or outcomes.
- Indirect – Important deviations from the protocol in at least two of the following areas: population, intervention, comparator and/or outcomes.

Methods for combining intervention evidence

Meta-analyses of interventional data were conducted with reference to the Cochrane Handbook for Systematic Reviews of Interventions (Higgins et al. 2011).

Where different studies presented continuous data measuring the same outcome but using different numerical scales (e.g. a 0-10 and a 0-100 visual analogue scale), these outcomes were all converted to the same scale before meta-analysis was conducted on the mean differences. Where outcomes measured the same underlying construct but used different instruments/metrics, data were analysed using standardised mean differences (Hedges' g).

A pooled relative risk was calculated for dichotomous outcomes (using the Mantel–Haenszel method). Both relative and absolute risks were presented, with absolute risks calculated by applying the relative risk to the pooled risk in the comparator arm of the meta-analysis.

Fixed- and random-effects models (der Simonian and Laird) were fitted for all syntheses, with the presented analysis dependent on the degree of heterogeneity in the assembled evidence. Fixed-effects models were the preferred choice to report, but in situations where the assumption of a shared mean for fixed-effects model were clearly not met, even after appropriate pre-specified subgroup analyses were conducted, random-effects results are presented. Fixed-effects models were deemed to be inappropriate if one or both of the following conditions was met:

- Significant between study heterogeneity in methodology, population, intervention or comparator was identified by the reviewer in advance of data analysis. This decision was made and recorded before any data analysis was undertaken.

- The presence of significant statistical heterogeneity in the meta-analysis, defined as $I^2 \geq 50\%$.

In any meta-analyses where some (but not all) of the data came from studies at high risk of bias, a sensitivity analysis was conducted, excluding those studies from the analysis. Results from both the full and restricted meta-analyses are reported. Similarly, in any meta-analyses where some (but not all) of the data came from indirect studies, a sensitivity analysis was conducted, excluding those studies from the analysis.

Meta-analyses were performed in Cochrane Review Manager v 5.3.

Minimal clinically important differences (MIDs)

The Core Outcome Measures in Effectiveness Trials (COMET) database was searched to identify published minimal clinically important difference thresholds relevant to this guideline. However, no relevant MIDs were found. In addition, the Guideline Committee were asked to specify any outcomes where they felt a consensus MID could be defined from their experience. The committee agreed that they could not specify any MIDs. Because all studies were cohort studies without a comparator, none of the studies had a line of no effect with which to rate imprecision. Therefore, imprecision was rated according to number of participants. If the number of participants in one arm was 40 or less, the committee agreed that the imprecision would most likely be serious. If the number of participants in one arm was 25 or less, the committee agreed that the imprecision would most likely be very serious.

GRADE for pairwise meta-analyses of interventional evidence

GRADE was used to assess the quality of evidence for the selected outcomes as specified in 'Developing NICE guidelines: the manual (2014)'. Data from RCTs was initially rated as high quality and the quality of the evidence for each outcome was downgraded or not from this initial point. If non-RCT evidence was included for intervention-type systematic reviews then these were initially rated as either moderate quality (quasi-randomised studies) or low quality (cohort studies) and the quality of the evidence for each outcome was further downgraded or not from this point, based on the criteria given in Table 4. The committee agreed that the outcomes of cohort studies with one arm (no comparator) would be described using a narrative synthesis.

Table 2: Rationale for downgrading quality of evidence for intervention studies

GRADE criteria	Reasons for downgrading quality
Risk of bias	Not serious: If less than 33.3% of the weight in a meta-analysis came from studies at moderate or high risk of bias, the overall outcome was not downgraded. Serious: If greater than 33.3% of the weight in a meta-analysis came from studies at moderate or high risk of bias, the outcome was downgraded one level. Very serious: If greater than 33.3% of the weight in a meta-analysis came from studies at high risk of bias, the outcome was downgraded two levels. Outcomes meeting the criteria for downgrading above were not downgraded if there was evidence the effect size was not meaningfully different between studies at high and low risk of bias.
Indirectness	Not serious: If less than 33.3% of the weight in a meta-analysis came from partially indirect or indirect studies, the overall outcome was not downgraded.

GRADE criteria	Reasons for downgrading quality
	Serious: If greater than 33.3% of the weight in a meta-analysis came from partially indirect or indirect studies, the outcome was downgraded one level. Very serious: If greater than 33.3% of the weight in a meta-analysis came from indirect studies, the outcome was downgraded two levels. Outcomes meeting the criteria for downgrading above were not downgraded if there was evidence the effect size was not meaningfully different between direct and indirect studies.
Inconsistency	Concerns about inconsistency of effects across studies, occurring when there is unexplained variability in the treatment effect demonstrated across studies (heterogeneity), after appropriate pre-specified subgroup analyses have been conducted. This was assessed using the I^2 statistic. N/A: Inconsistency was marked as not applicable if data on the outcome was only available from one study. Not serious: If the I^2 was less than 33.3%, the outcome was not downgraded. Serious: If the I^2 was between 33.3% and 66.7%, the outcome was downgraded one level. Very serious: If the I^2 was greater than 66.7%, the outcome was downgraded two levels. Outcomes meeting the criteria for downgrading above were not downgraded if there was evidence the effect size was not meaningfully different between studies with the smallest and largest effect sizes.
Imprecision	The line of no effect was defined as the MID. The outcome was downgraded once if the 95% confidence interval for the effect size crossed the line of no effect (i.e. the outcome was not statistically significant), and twice if the sample size of the study was sufficiently small that it is not plausible any realistic effect size could have been detected. The committee agreed that a sample size of 25 or less would result in such a downgrade. Outcomes meeting the criteria for downgrading above were not downgraded if the confidence interval was sufficiently narrow that the upper and lower bounds would correspond to clinically equivalent scenarios.

The quality of evidence for each outcome was upgraded if any of the following five conditions were met:

- Data from non-randomised studies showing an effect size sufficiently large that it cannot be explained by confounding alone.
- Data showing a dose-response gradient.
- Data where all plausible residual confounding is likely to increase our confidence in the effect estimate.

Publication bias

Publication bias was assessed in two ways. First, if evidence of conducted but unpublished studies was identified during the review (e.g. conference abstracts, trial protocols or trial records without accompanying published data), available information on these unpublished studies was reported as part of the review. Secondly, where 10 or more studies were included as part of a single meta-analysis, a funnel plot was produced to graphically assess the potential for publication bias.

Evidence statements

Evidence statements for pairwise intervention data are classified in to one of four categories:

- Situations where the data are only consistent, at a 95% confidence level, with an effect in one direction (i.e. one that is 'statistically significant'), and the magnitude of that effect is most likely to meet or exceed the MID (i.e. the point estimate is not in the zone of equivalence). In such cases, we state that the evidence showed that there is an effect.
- Situations where the data are only consistent, at a 95% confidence level, with an effect in one direction (i.e. one that is 'statistically significant'), but the magnitude of that effect is most likely to be less than the MID (i.e. the point estimate is in the zone of equivalence). In such cases, we state that the evidence could not demonstrate a meaningful difference.
- Situations where the data are consistent, at a 95% confidence level, with an effect in either direction (i.e. one that is not 'statistically significant') but the confidence limits are smaller than the MIDs in both directions. In such cases, we state that the evidence demonstrates that there is no difference.
- In all other cases, we state that the evidence could not differentiate between the comparators.

Diagnostic test accuracy evidence

In this guideline, diagnostic test accuracy (DTA) data are classified as any data in which a test result or the output of an algorithm – is observed in some people who have the condition of interest at the time of the test and some people who do not. Such data either explicitly provide, or can be manipulated to generate, a 2x2 classification of true positives and false negatives (in people who, according to the reference standard, truly have the condition) and false positives and true negatives (in people who, according to the reference standard, do not).

The 'raw' 2x2 data can be summarised in a variety of ways. Those that were used for decision making in this guideline are as follows:

- **Sensitivity** is the probability that the feature will be positive in a person with the condition.
  - sensitivity = TP/(TP+FN)
- **Specificity** is the probability that the feature will be negative in a person without the condition.
  - specificity = TN/(TN+FP)

Meta-analysis of diagnostic test accuracy was undertaken for this guideline using univariate random effects models, which were effectively four simple meta-analyses of a proportion. We were unable to fit a bivariate model due to having a small number of studies for both CT and MRI. Bayesian methods were chosen in order to handle zero-cells without the need for a continuity correction with vague prior distributions being assigned to sensitivity and specificity for the two tests. Random effects models were preferred based on DIC being more than 3-5 points lower for sensitivity and because of heterogeneity in study populations, methods and settings. While the DIC for the random effects model for specificity was not 3-5 points lower, it was still preferred due to heterogeneity in study populations, methods and settings. Further details can be found in Appendix F (GRADE tables), Appendix I (Cost-utility analysis) and Appendix K (WinBUGS code).

Quality assessment

Individual studies were quality assessed using the QUADAS-2 tool, which contains four domains: patient selection, index test, reference standard, and flow and timing. Each individual study was classified into one of the following two groups:

- Low risk of bias – Evidence of non-serious bias in zero or one domain.
- Moderate risk of bias – Evidence of non-serious bias in two domains only, or serious bias in one domain only.
- High risk of bias – Evidence of bias in at least three domains, or of serious bias in at least two domains.

Each individual study was also classified into one of three groups for directness, based on if there were concerns about the population, index features and/or reference standard in the study and how directly these variables could address the specified review question. Studies were rated as follows:

- Direct – No important deviations from the protocol in population, index feature and/or reference standard.
- Partially indirect – Important deviations from the protocol in one of the population, index feature and/or reference standard.
- Indirect – Important deviations from the protocol in at least two of the population, index feature and/or reference standard.

Modified GRADE for diagnostic test accuracy evidence

GRADE has not been developed for use with diagnostic studies; therefore a modified approach was applied using the GRADE framework. GRADE assessments were only undertaken for sensitivity and specificity. This is because the committee agreed that these two measurements are the ones that that matter most to clinicians and people with stage I to stage IIIA lung cancer being considered for radical treatment. GRADE quality ratings were calculated using the same criteria as for intervention studies, given in Table 4. Neither sensitivity nor specificity have a line of no effect with which to rate imprecision. Therefore, imprecision was rated according to number of participants. If the number of participants in one arm was 40 or less, the committee agreed that the imprecision would most likely be serious. If the number of participants in one arm was 25 or less, the committee agreed that the imprecision would most likely be very serious.

Appendix C – Literature search strategies

Scoping search strategies

Scoping searches Scoping searches were undertaken on the following websites and databases (listed in alphabetical order) in April 2017 to provide information for scope development and project planning. Browsing or simple search strategies were employed.

Guidelines/website
American Cancer Society
American College of Chest Physicians
American Society for Radiation Oncology
American Thoracic Society
Association for Molecular Pathology
British Lung Foundation
British Thoracic Society
Canadian Medical Association Infobase
Canadian Task Force on Preventive Health Care
Cancer Australia
Cancer Care Ontario
Cancer Control Alberta
Cancer Research UK
Care Quality Commission
College of American Pathologists
Core Outcome Measures in Effectiveness Trials (COMET)
Department of Health & Social Care
European Respiratory Society
European Society for Medical Oncology
European Society of Gastrointestinal Endoscopy
European Society of Thoracic Surgery
General Medical Council
Guidelines & Audit Implementation Network (GAIN)
Guidelines International Network (GIN)
Healthtalk Online
International Association for the Study of Lung Cancer
MacMillan Cancer Support
Medicines and Products Regulatory Agency (MHRA)
National Audit Office
National Cancer Intelligence Network
National Clinical Audit and Patient Outcomes Programme
National Health and Medical Research Council - Australia
National Institute for Health and Care Excellence (NICE) - published & in development guidelines
National Institute for Health and Care Excellence (NICE) - Topic Selection
NHS Choices
NHS Digital
NHS England

Guidelines/website
NICE Clinical Knowledge Summaries (CKS)
NICE Evidence Search
Office for National Statistics
Patient UK
PatientVoices
Public Health England
Quality Health
Royal College of Anaesthetists
Royal College of General Practitioners
Royal College of Midwives
Royal College of Nursing
Royal College of Pathologists
Royal College of Physicians
Royal College of Radiologists
Royal College of Surgeons
Scottish Government
Scottish Intercollegiate Guidelines Network (SIGN)
UK Data Service
US National Guideline Clearinghouse
Walsall community Health NHS Trust
Welsh Government

Clinical search literature search strategy

Main searches

Bibliographic databases searched for the guideline

- Cochrane Database of Systematic Reviews – CDSR (Wiley)
- Cochrane Central Register of Controlled Trials – CENTRAL (Wiley)
- Database of Abstracts of Reviews of Effects – DARE (Wiley)
- Health Technology Assessment Database – HTA (Wiley)
- EMBASE (Ovid)
- MEDLINE (Ovid)
- MEDLINE Epub Ahead of Print (Ovid)
- MEDLINE In-Process (Ovid)

Identification of evidence for review questions

The searches were conducted between October 2017 and April 2018 for 9 review questions (RQ).

Searches were re-run in May 2018.

Where appropriate, in-house study design filters were used to limit the retrieval to, for example, randomised controlled trials. Details of the study design filters used can be found in section 3.

Search strategy

Medline Strategy, searched 13th February 2018
Database: Ovid MEDLINE(R) 1946 to Present with Daily Update
Search Strategy:

1 Small Cell Lung Carcinoma/
2 Carcinoma, Small Cell/
3 SCLC.tw.
4 ((pancoast* or superior sulcus or pulmonary sulcus) adj4 (tumo?r* or syndrome*)).tw.
5 or/1-4
6 ((small or oat or reserve or round) adj1 cell adj1 (lung* or pulmonary or bronch*) adj3 (cancer* or neoplasm* or carcinoma* or tumo?r* or lymphoma* or metast* or malignan* or blastoma* or carcinogen* or adenocarcinoma* or angiosarcoma* or chrondosarcoma* or sarcoma* or teratoma* or microcytic*)).tw.
7 (non adj1 small adj1 cell adj1 (lung* or pulmonary or bronch*) adj3 (cancer* or neoplasm* or carcinoma* or tumo?r* or lymphoma* or metast* or malignan* or blastoma* or carcinogen* or adenocarcinoma* or angiosarcoma* or chrondosarcoma* or sarcoma* or teratoma* or microcytic*)).tw.
8 6 not 7
9 5 or 8
10 exp Radiotherapy/
11 Radiation Oncology/
12 exp Radiography, Thoracic/
13 radiotherapy.fs.
14 (radiotherap* or radiotreat* or roentgentherap* or radiosurg*).tw.
15 ((radiat* or radio* or irradiat* or roentgen or x-ray or xray) adj4 (therap* or treat* or repair* or oncolog* or surg*)).tw.
16 (RT or RTx or XRT or TRT or TCRT).tw.
17 or/10-16
18 9 and 17
19 limit 18 to english language
20 Animals/ not Humans/
21 19 not 20

Study Design Filters

The MEDLINE SR, RCT, and observational studies filters are presented below.

Systematic Review

1. Meta-Analysis.pt.
2. Meta-Analysis as Topic/
3. Review.pt.
4. exp Review Literature as Topic/
5. (metaanaly$ or metanaly$ or (meta adj3 analy$)).tw.
6. (review$ or overview$).ti.
7. (systematic$ adj5 (review$ or overview$)).tw.
8. ((quantitative$ or qualitative$) adj5 (review$ or overview$)).tw.
9. ((studies or trial$) adj2 (review$ or overview$)).tw.
10. (integrat$ adj3 (research or review$ or literature)).tw.
11. (pool$ adj2 (analy$ or data)).tw.
12. (handsearch$ or (hand adj3 search$)).tw.

The MEDLINE SR, RCT, and observational studies filters are presented below.
13. (manual$ adj3 search$).tw. 14. or/1-13 15. animals/ not humans/ 16. 14 not 15
RCT 1 Randomized Controlled Trial.pt. 2 Controlled Clinical Trial.pt. 3 Clinical Trial.pt. 4 exp Clinical Trials as Topic/ 5 Placebos/ 6 Random Allocation/ 7 Double-Blind Method/ 8 Single-Blind Method/ 9 Cross-Over Studies/ 10 ((random$ or control$ or clinical$) adj3 (trial$ or stud$)).tw. 11 (random$ adj3 allocat$).tw. 12 placebo$.tw. 13 ((singl$ or doubl$ or trebl$ or tripl$) adj (blind$ or mask$)).tw. 14 (crossover$ or (cross adj over$)).tw. 15 or/1-14 16 animals/ not humans/ 17 15 not 16
Observational 1 Observational Studies as Topic/ 2 Observational Study/ 3 Epidemiologic Studies/ 4 exp Case-Control Studies/ 5 exp Cohort Studies/ 6 Cross-Sectional Studies/ 7 Controlled Before-After Studies/ 8 Historically Controlled Study/ 9 Interrupted Time Series Analysis/ 10 Comparative Study.pt. 11 case control$.tw. 12 case series.tw. 13 (cohort adj (study or studies)).tw. 14 cohort analy$.tw. 15 (follow up adj (study or studies)).tw. 16 (observational adj (study or studies)).tw. 17 longitudinal.tw. 18 prospective.tw. 19 retrospective.tw. 20 cross sectional.tw. 21 or/1-20

Health Economics literature search strategy

Sources searched to identify economic evaluations

- NHS Economic Evaluation Database – NHS EED (Wiley) last updated Apr 2015

- Health Technology Assessment Database – HTA (Wiley) last updated Oct 2016
- Embase (Ovid)
- MEDLINE (Ovid)
- MEDLINE In-Process (Ovid)

Search filters to retrieve economic evaluations and quality of life papers were appended to the review question search strategies. For some health economics strategies additional terms were added to the original review question search strategies (see sections 4.2, 4.3 and 4.4) The searches were conducted between October 2017 and April 2018 for 9 review questions (RQ).

Searches were re-run in May 2018.

Searches were limited to those in the English language. Animal studies were removed from results.

Economic evaluation and quality of life filters

Medline Strategy
Economic evaluations
1 Economics/
2 exp "Costs and Cost Analysis"/
3 Economics, Dental/
4 exp Economics, Hospital/
5 exp Economics, Medical/
6 Economics, Nursing/
7 Economics, Pharmaceutical/
8 Budgets/
9 exp Models, Economic/
10 Markov Chains/
11 Monte Carlo Method/
12 Decision Trees/
13 econom$.tw.
14 cba.tw.
15 cea.tw.
16 cua.tw.
17 markov$.tw.
18 (monte adj carlo).tw.
19 (decision adj3 (tree$ or analys$)).tw.
20 (cost or costs or costing$ or costly or costed).tw.
21 (price$ or pricing$).tw.
22 budget$.tw.
23 expenditure$.tw.
24 (value adj3 (money or monetary)).tw.
25 (pharmacoeconomic$ or (pharmaco adj economic$)).tw.
26 or/1-25
Quality of life
1 "Quality of Life"/
2 quality of life.tw.

Medline Strategy
3 "Value of Life"/
4 Quality-Adjusted Life Years/
5 quality adjusted life.tw.
6 (qaly$ or qald$ or qale$ or qtime$).tw.
7 disability adjusted life.tw.
8 daly$.tw.
9 Health Status Indicators/
10 (sf36 or sf 36 or short form 36 or shortform 36 or sf thirtysix or sf thirty six or shortform thirtysix or shortform thirty six or short form thirtysix or short form thirty six).tw.
11 (sf6 or sf 6 or short form 6 or shortform 6 or sf six or sfsix or shortform six or short form six).tw.
12 (sf12 or sf 12 or short form 12 or shortform 12 or sf twelve or sftwelve or shortform twelve or short form twelve).tw.
13 (sf16 or sf 16 or short form 16 or shortform 16 or sf sixteen or sfsixteen or shortform sixteen or short form sixteen).tw.
14 (sf20 or sf 20 or short form 20 or shortform 20 or sf twenty or sftwenty or shortform twenty or short form twenty).tw.
15 (euroqol or euro qol or eq5d or eq 5d).tw.
16 (qol or hql or hqol or hrqol).tw.
17 (hye or hyes).tw.
18 health$ year$ equivalent$.tw.
19 utilit$.tw.
20 (hui or hui1 or hui2 or hui3).tw.
21 disutili$.tw.
22 rosser.tw.
23 quality of wellbeing.tw.
24 quality of well-being.tw.
25 qwb.tw.
26 willingness to pay.tw.
27 standard gamble$.tw.
28 time trade off.tw.
29 time tradeoff.tw.
30 tto.tw.
31 or/1-30

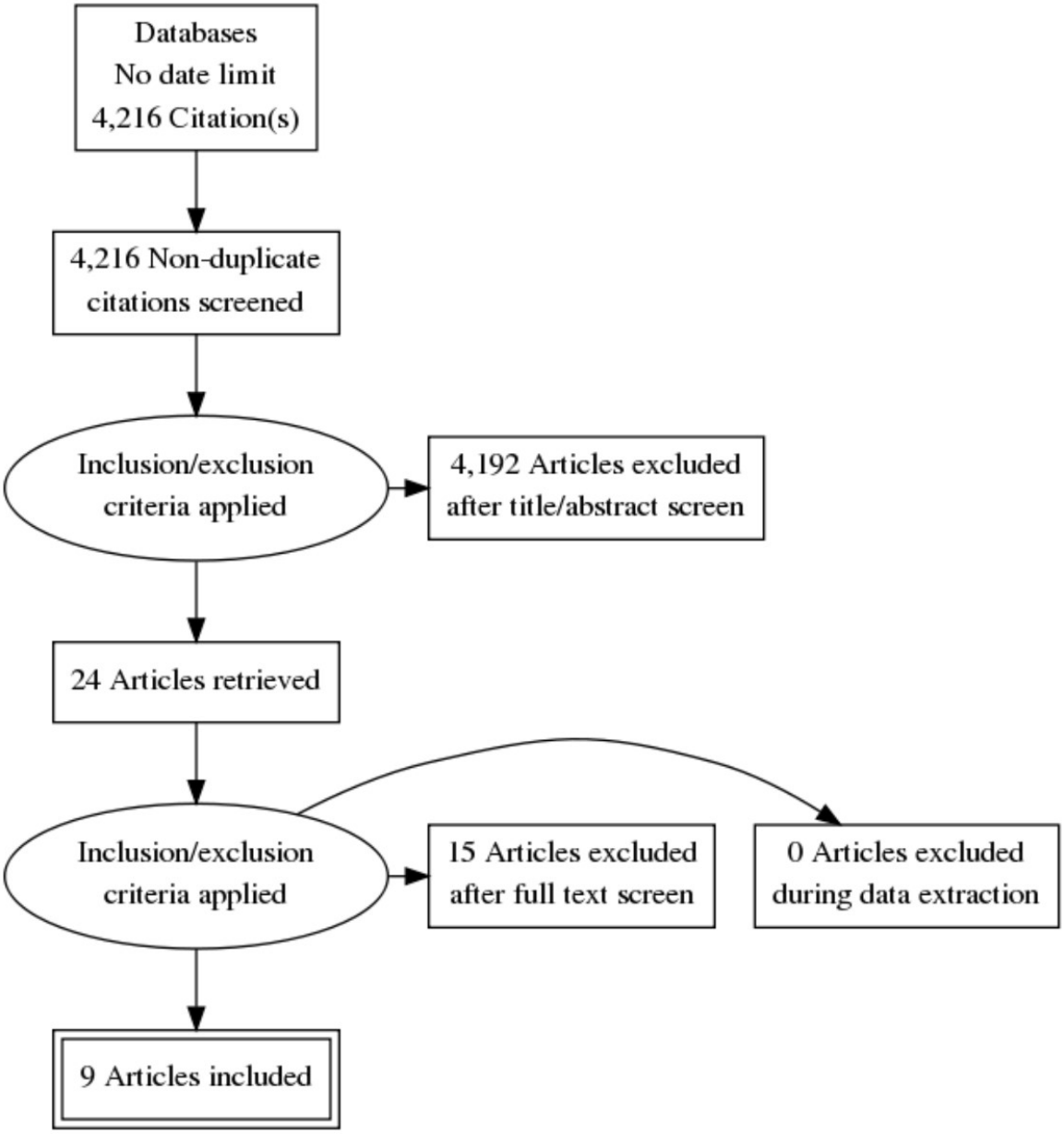
Health economics search strategy

Medline Strategy, searched 5th December 2017 Database: Ovid MEDLINE(R) 1946 to Present with Daily Update Search Strategy:
1 exp Lung Neoplasms/
2 ((lung* or pulmonary or bronch*) adj3 (cancer* or neoplasm* or carcinoma* or tumo?r* or lymphoma* or metast* or malignan* or blastoma* or carcinogen* or adenocarcinoma* or angiosarcoma* or chrondosarcoma* or sarcoma* or teratoma* or microcytic*)).tw.
3 ((pancoast* or superior sulcus or pulmonary sulcus) adj4 (tumo?r* or syndrome*)).tw.
4 ((lung* or pulmonary or bronch*) adj4 (oat or small or non-small) adj4 cell*).tw.
5 (SCLC or NSCLC).tw.
6 or/1-5
7 exp Magnetic Resonance Imaging/

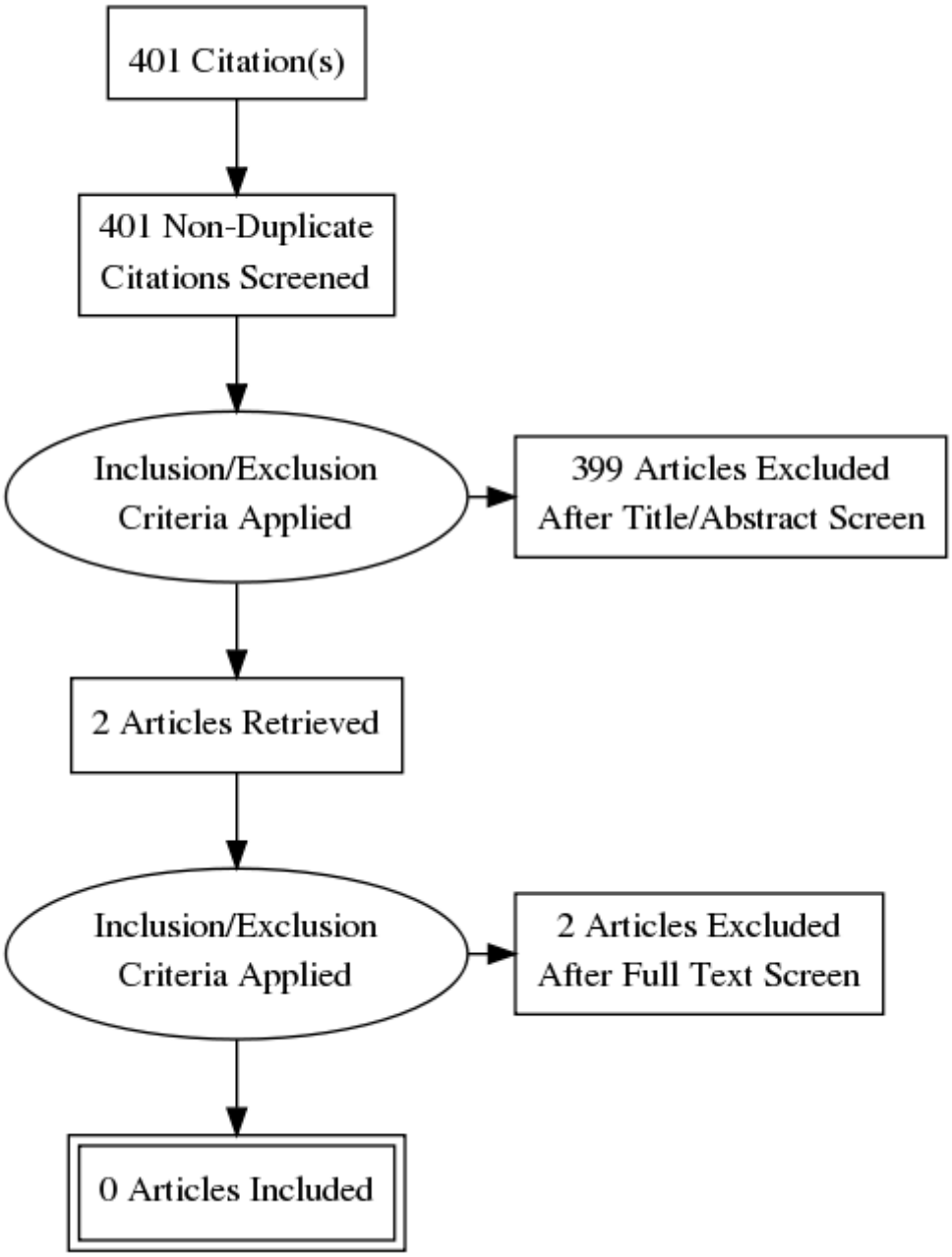
Medline Strategy, searched 5th December 2017 **Database: Ovid MEDLINE(R) 1946 to Present with Daily Update** **Search Strategy:**
8 ((magnet* or spin* or chemical* or proton*) adj4 (resonan* or shift* or spin* or echo* or contrast* or transfer*) adj4 (imag* or tomograph* or angiograph* or perfusion*)).tw.
9 ((NMR or MR) adj4 (imag* or tomograph* or angiograph* or perfusion*)).tw.
10 (MRI or MRIs or fMRI or MRA or DWI).tw.
11 zeugmatograph*.tw.
12 exp Magnetic Resonance Spectroscopy/
13 (magnet* adj4 resonan* adj4 (spectro* or spectra* or spectru*)).tw.
14 ((NMR or MR) adj4 (spectro* or spectra* or spectru*)).tw.
15 MRS.tw.
16 (fluid attenuated adj4 inversion recover*).tw.
17 FLAIR.tw.
18 exp Tomography, X-Ray Computed/
19 Tomography Scanners, X-Ray Computed/
20 ((comput* or electron*) adj4 tomograph*).tw.
21 (CT or CAT or NCCT or MDCT or PCT).tw.
22 tomodensitometr*.tw.
23 or/7-22
24 exp Central Nervous System/
25 exp Skull/
26 (brain* or encephalon* or cerebr* or intracranial* or supratentorial* or cerebell* or mening* or leptomening*).tw.
27 ((gr?y or white) adj2 matter*).tw.
28 (cranial* or skull* or cranium* or calvari* or pituitar* or hypophys* or infundibl* or infracerebral*).tw.
29 (central nervous system* or CNS or cerebrospin*).tw. (274289)
30 or/24-29
31 23 and 30
32 exp Neuroimaging/
33 neuroimag*.tw.
34 ((brain* or encephalon* or cerebr* or intracranial* or supratentorial* or cerebell* or mening* or leptomening* or neuro* or neura*) adj4 (imag* or scan* or tomograph*)).tw.
35 exp Central Nervous System/dg [Diagnostic Imaging]
36 exp Brain Neoplasms/dg [Diagnostic Imaging]
37 Cerebrospinal Fluid/dg [Diagnostic Imaging]
38 or/32-37
39 31 or 38
40 6 and 39
41 Animals/ not Humans/
42 40 not 41

Appendix D – Evidence study selection

Clinical Evidence study selection



Economic Evidence study selection



Appendix E – Clinical evidence tables

Short Title	Title	Study Characteristics	Risk of Bias: quality assessment
de Cos Escuin 2007	Silent brain metastasis in the initial staging of lung cancer: evaluation by computed tomography and magnetic resonance imaging	**Study type** • Retrospective cohort study **Study details** • Study location *Spain* • Study setting *Hospital* • Study dates *2000 to 2005* • Duration of follow-up *There were 4 participants who had no brain metastases on the initial imaging but who later went on to have brain metastases. Three of these occurred at in under 3 months, the fourth occurred at 17 months. The latter case was not included in the analysis because the time delay meant that it could be a new metastasis that seeded after the initial brain imaging. The method of follow-up was described as 'routine' but no further details were given.* • Loss to follow-up *None* • Sources of funding **Not mentioned** **Inclusion criteria** • Histopathologically proven lung cancer **Exclusion criteria**	**Quality assessment (cohort study)** Did the study address a clearly focused issue? • Yes Was the cohort recruited in an acceptable way? • No *There is no explanation with regards to decision making as to who got an MRI brain and who got a CT brain. In the MRI group, there were 26 people who were stage I or II compared to 12 people in the CT group who were stage I or II.* Was the exposure accurately measured to minimise bias? • Yes Was the outcome accurately measured to minimise bias? • No *Time is not a recommended 'gold standard' because brain metastases could have been seeded after the screening brain MRI. However, we appreciate the difficulty in selecting a better gold standard.* Have the authors identified all-important confounding factors? • Yes

Short Title	Title	Study Characteristics	Risk of Bias: quality assessment
		• Neurologic symptoms and signs that were suggestive of brain metastasis **Sample characteristics** • Sample size *173 people* • Split between study groups *MRI brain = 97; CT brain = 76* • Loss to follow-up *None* • %female *MRI brain = 3%; CT = 11%* • Average age *Mean (range): MRI brain = 67.8 years (37-88); CT brain = 67.5 years (38-82)* **Index test / intervention (first arm of study)** • MRI brain *Cranial MRI was performed on a 1.5-T NT Gyroscan scanner. T1-weighted precontrast images (repetition time of 600 ms and echo time of 17 ms) and T2-weighted images (repetition time of 4900 ms and echo time of 120 ms) were acquired. The field of vision was 20 cm × 20 cm and the matrix 256 mm × 256 mm. Section thickness was 6 mm, with a 1-mm intersection gap. The T1-weighted images were repeated following the administration of 0.2 mL/kg of a paramagnetic gadolinium-based contrast agent.* **Intervention 2 (second arm of study)** • CT brain	Have they taken account of the confounding factors in the design and/or analysis? • No *The higher the lung cancer grade, the greater the chances of no surgery. This might erroneously produce false negatives.* Was the follow up of subjects complete enough? • Yes Was the follow up of subjects long enough? • Yes Overall risk of bias • High *69% of the CT participants were stage IIIB and above, 55% of the MRI participants were stage IIIB and above: 8 of these developed brain metastases. These participants would not have had surgery to remove cancer. Therefore, there might be a greater possibility of metastases seeding to the brain after the brain imaging, producing false negatives in error.* Directness • Indirectly applicable *69% of the CT participants were stage IIIB and above, 55% of the MRI participants were stage IIIB and above: 8 of these developed brain metastases. Therefore, 6 developed brain metastases in the population of interest. Unfortunately, we do not know what arms they were in.*

Short Title	Title	Study Characteristics	Risk of Bias: quality assessment
		The cranial CT study was performed on a Tomoscan AV scanner using contiguous 5-mm to 10-mm slices, and the images were contrast enhanced with 50 mL of iopromide. **Reference standard** • Follow-up: new brain metastases were discounted at a time that was decided during the analysis *There were 4 participants who had no brain metastases on the initial imaging but who later went on to have brain metastases. Three of these occurred at in under 3 months, the fourth occurred at 17 months. The latter case was not included in the analysis because the time delay meant that it could be a new metastasis that seeded after the initial brain imaging.* **Outcomes (study was part diagnostic, part intervention)** • Diagnostic sensitivity and specificity • Change in treatment plan: brain metastases discovered using MRI brain	**Quality assessment (diagnostic test accuracy review – QUADAS 2)** Was a consecutive or random sample of patients enrolled? • No *Consecutive* Was a case-control design avoided? • Yes Did the study avoid inappropriate exclusions? • Yes RISK Could the selection of patients have introduced bias? • High *There is no explanation with regards to decision making as to who got an MRI brain and who got a CT brain. In the MRI group, there were 26 people who were stage I or II compared to 12 people in the CT group who were stage I or II.* CONCERN Is there concern that the included patients do not match the review question? • High *69% of the CT participants were stage IIIB and above, 55% of the MRI participants were stage IIIB and above: 8 of these developed brain metastases. Therefore, 6 developed brain metastases in the population of interest. Unfortunately, we do not know what arms they were in.*

Short Title	Title	Study Characteristics	Risk of Bias: quality assessment
			Were the index test results interpreted without knowledge of the results of the reference standard? • No *No blinding* If a threshold was used, was it pre-specified? • Yes RISK Could the conduct or interpretation of the index test have introduced bias? • Low Concerns regarding applicability • Low Is the reference standard likely to correctly classify the target condition? • Yes Were the reference standard results interpreted without knowledge of the results of the index test? • No *No blinding* RISK Could the reference standard, its conduct, or its interpretation have introduced bias? • High *Time is not a recommended 'gold standard' because brain metastases could have been seeded after the*

Short Title	Title	Study Characteristics	Risk of Bias: quality assessment
			screening brain MRI. However, we appreciate the difficulty in selecting a better gold standard. CONCERN Is there concern that the target condition as defined by the reference standard does not match the review question? • High *69% of the CT participants were stage IIIB and above, 55% of the MRI participants were stage IIIB and above: 8 of these developed brain metastases. Therefore, 6 developed brain metastases in the population of interest. Unfortunately, we do not know what arms they were in.* Was there an appropriate interval between index test(s) and reference standard? • No *Time is being used as the gold standard. Metastases could be seeded after the brain imaging.* Did all patients receive a reference standard? • Yes Did patients receive the same reference standard? • Yes Were all patients included in the analysis? • Yes RISK Could the patient flow have introduced bias? • Low

Short Title	Title	Study Characteristics	Risk of Bias: quality assessment
			Overall quality • Low
Earnest 1999	Suspected non-small cell lung cancer: incidence of occult brain and skeletal metastases and effectiveness of imaging for detection--pilot study	**Study type** • Prospective cohort study **Study details** • Study location *USA* • Study setting *Hospital (Mayo Clinic)* • Study dates *Not provided. This study was published in 1999.* • Duration of follow-up *12 months* • Loss to follow-up *Complete follow-up information through 12 months was available for 26 (90%) of 29 study patients (25 of 27 with NSCLC). Of the three study group patients who did not complete follow-up, one underwent resection of a lung metastasis from colon cancer, another had insufficient pulmonary reserve for pneumonectomy and requested to be removed from the study after 9 months of follow-up, and a third was unable to complete preoperative MRI due to claustrophobia; this third patient underwent left upper lobectomy for stage IB (T2N0M0) squamous cell carcinoma and did not complete questionnaires beyond 6 months of follow-up.* • Sources of funding *Mayo Foundation for Education and Research, and Bracco Diagnostics. Bracco Diagnostics makes CT and MRI contrast agents.* **Inclusion criteria**	**Quality assessment (cohort study)** Did the study address a clearly focused issue? • Yes Was the cohort recruited in an acceptable way? • No *There is no explanation or method to how people were assigned to the MRI brain arm or to the control arm. The MRI brain group had 48% of people with squamous cell carcinoma and 34.5% of people with adenocarcinoma. The comparison group had 41% of people with squamous cell carcinoma and 47.3% of people with adenocarcinoma. In addition, the MRI brain group had 2/29 who were N3 and the comparison group had 0/110 who were N3. Therefore, the groups were not balanced.* Was the exposure accurately measured to minimise bias? • Yes Was the outcome accurately measured to minimise bias? • No *No outcomes were looked at in the comparison group. Time is a poor reference standard because metastases could have been seeded after the initial brain imaging.*

Short Title	Title	Study Characteristics	Risk of Bias: quality assessment
		• Staging CT chest & abdomen • Suspected of having lung cancer *Lung cancer was confirmed later using histology.* **Exclusion criteria** • Unable to undergo brain MRI, for example: cardiac pacemaker, cochlear implant, intracranial aneurysm clip, known sensitivity to MRI contrast agents, presence of renal failure • Lung mass 3 cm or smaller • Clinical evidence of remote metastases • Evidence of mediastinal invasion • Evidence of abdominal metastases • Pregnancy or lactation • History of lung, breast or renal cancer in the last 5 years • Cardiac pacemaker • Inability to tolerate a curative surgical procedure **Sample characteristics** • Sample size *MRI brain group = 28; no MRI brain group = 110* • %female *MRI brain group = 24.1%; no MRI brain group = 36.4%* • Average age *Median age (range): MRI brain group = 67.2 years (48-80); no MRI brain group = 71.3 years (24-86)* **Index test / intervention (first arm of study)** • MRI brain *MRI was performed with a 1.5-T imager. MRI included sagittal T1-weighted spin-echo imaging (repetition time msec/echo time msec,*	Have the authors identified all important confounding factors? • Yes Have they taken account of the confounding factors in the design and/or analysis? • Yes Was the follow up of subjects complete enough? • Yes Was the follow up of subjects long enough? • Yes Overall risk of bias • High Directness • Partially directly applicable *People with lung cancer T1 were excluded (mass 3 cm or less). Our inclusion criteria are stages I to IIIA. The MRI brain images were reviewed by 2 radiologists. In clinical practice, 1 radiologist might review the images. In the study, 1 radiologist made a false negative error (that was corrected by the second radiologist).* **Quality assessment (diagnostic test accuracy review – QUADAS 2)**

Short Title	Title	Study Characteristics	Risk of Bias: quality assessment
		300/15; two signals acquired, 512 3 512 matrix, 48-cm field of view, 3-mm section thickness, 0.5-mm intersection gap), performed with a phased-array coil, of the cervical, thoracic, and lumbosacral spine and coronal T1-weighted spin-echo imaging (350/16, two signals acquired 256 3 192 matrix, 48-cm field of view, 5-mm interleaved section thickness), performed with a body coil, of the pelvis and proximal portions of the femurs. MRI of the brain included sagittal T1-weighted spin-echo imaging (600/16, two signals acquired, 256 3 192 matrix, 24-cm field of view, 5-mm section thickness) performed prior to contrast agent administration, axial T2-weighted spin-echo imaging (2,500/ 30–90, three-fourths signal acquired, 256 3 192 matrix, 20-cm field of view, 5-mm section thickness, 2.5-mm intersection gap) performed immediately after administration of a conventional dose of 0.1 mmol/kg of gadoteridol followed by axial T1-weighted spin-echo imaging (450/16, two signals acquired, 256 3 192 matrix, 20-cm field of view, 5-mm section thickness, 1-mm intersection gap) performed at least 10 minutes after contrast agent administration. A second dose of gadoteridol was then administered, for a total dose of 0.3 mmol/kg of gadoteridol (high dose), which was immediately followed by acquisition of a second series of axial T1-weighted spin-echo images with the same parameters. Two neuroradiologists reviewed the MRI images. The reviewing radiologists were informed that the studies had been obtained in patients suspected of having lung cancer. **Reference standard** • Follow-up for 12 months *Questionnaires were sent to each patient every 3 months after study entry to determine the incidence of clinical metastatic disease to the brain. Any follow-up imaging studies of the brain in the 29 study patients were obtained and reviewed, if possible. Confirmation of metastatic disease to the brain was established by means of biopsy and/or resection results or progressive lesion enlargement demonstrated on successive follow-up CT or MR studies. Negative preoperative imaging studies were considered to be false-negative if a*	Was a consecutive or random sample of patients enrolled? • Unclear *Not mentioned* Was a case-control design avoided? • Yes Did the study avoid inappropriate exclusions? • Yes RISK Could the selection of patients have introduced bias? • High *There is no explanation or method to how people were assigned to the MRI brain arm or to the control arm. The MRI brain group had 48% of people with squamous cell carcinoma and 34.5% of people with adenocarcinoma. The comparison group had 41% of people with squamous cell carcinoma and 47.3% of people with adenocarcinoma. In addition, the MRI brain group had 2/29 who were N3 and the comparison group had 0/110 who were N3. Therefore, the groups were not balanced. Relatively small number of participants* CONCERN Is there concern that the included patients do not match the review question? • Low Were the index test results interpreted without knowledge of the results of the reference standard?

Short Title	Title	Study Characteristics	Risk of Bias: quality assessment
		metastatic lesion was detected in the 12 months of follow-up and if the lesion was documented with subsequent imaging studies. Likewise, positive preoperative imaging studies were considered to be false-positive in the absence of histologic proof or if lesion stability or resolution of an imaging abnormality was demonstrated during the 12-month follow-up. **Outcomes (study was part diagnostic, part intervention)** • Diagnostic sensitivity and specificity • Change in treatment plan: brain metastases discovered using MRI brain	• Yes If a threshold was used, was it pre-specified? • Yes RISK Could the conduct or interpretation of the index test have introduced bias? • High Concerns regarding applicability • Low Is the reference standard likely to correctly classify the target condition? • Unclear *The amount of follow-up time needed is not known.* Were the reference standard results interpreted without knowledge of the results of the index test? • No RISK Could the reference standard, its conduct, or its interpretation have introduced bias? • High CONCERN Is there concern that the target condition as defined by the reference standard does not match the review question? • Low

Short Title	Title	Study Characteristics	Risk of Bias: quality assessment
			Was there an appropriate interval between index test(s) and reference standard? • No *Reference standard tests are normally done at a similar time to the index test.* Did all patients receive a reference standard? • Yes Did patients receive the same reference standard? • Yes Were all patients included in the analysis? • Yes RISK Could the patient flow have introduced bias? • High Overall quality • Low
Ferrigno 1994	Cranial computed tomography as a part of the initial staging procedures for patients with non-small-cell lung cancer	**Study type** • Retrospective cohort study **Study details** • Study location *Italy* • Study setting *Hospital* • Study dates *1988 to 1991*	**Quality assessment (cohort study)** Did the study address a clearly focused issue? • Yes Was the cohort recruited in an acceptable way? • Yes Was the exposure accurately measured to minimise bias? • No

Short Title	Title	Study Characteristics	Risk of Bias: quality assessment
		• Duration of follow-up *12 months. However, the details of how participants were followed up is not provided. For example, there is no description of scheduled follow-ups at a specific time.* • Loss to follow-up *This could have been high because there are no details of specific follow-up period(s).* • Sources of funding *Not mentioned* **Inclusion criteria** • Histopathologically proven lung cancer **Exclusion criteria** • None mentioned **Sample characteristics** • Sample size *184 people* • Split between study groups *N/A* • Loss to follow-up *Not mentioned. This could have been high because there are no details of specific follow-up period(s).* • %female *9.2%* • Average age *Median (range): 63 years (41-85)* **Index test / intervention (first arm of study)**	*63% of participants are stage IIIB and above. Most, if not all of these participants would not have been operated on. As a result, we might expect a higher probability of metastases seeding in the brain after the initial brain imaging. This might produce a greater number of false negative results in error.* Was the outcome accurately measured to minimise bias? • No *Time is a poor reference standard because metastases could have been seeded after the initial brain imaging.* Have the authors identified all important confounding factors? • No *63% of participants are stage IIIB and above. Most, if not all of these participants would not have been operated on. As a result, we might expect a higher probability of metastases seeding in the brain after the initial brain imaging. This might produce a greater number of false negative results in error.* Have they taken account of the confounding factors in the design and/or analysis? • No *63% of participants are stage IIIB and above. Most, if not all of these participants would not have been operated on. As a result, we might expect a higher probability of metastases seeding in the brain after*

Short Title	Title	Study Characteristics	Risk of Bias: quality assessment
		• CT brain *All CT scans were performed on a scanner for brain CT, 50 ml non-iodinated intravenous contrast was injected prior to all studies.* **Reference standard** • Follow-up for 12 months *However, the details of how participants were followed up is not provided. For example, there is no description of scheduled follow-ups at a specific time.* **Outcomes (study was part diagnostic, part intervention)** • Diagnostic sensitivity and specificity • Change in treatment plan: brain metastases discovered using MRI brain	*the initial brain imaging. This might produce a greater number of false negative results in error.* Was the follow up of subjects complete enough? • No *The drop-out rate could have been high because there are no details of specific follow-up period(s).* Was the follow up of subjects long enough? • Yes Overall risk of bias • High Directness • Indirectly applicable 63% of participants were stage IIIB or above. **Quality assessment (diagnostic test accuracy review – QUADAS 2)** Was a consecutive or random sample of patients enrolled? • No *Consecutive* Was a case-control design avoided? • Yes Did the study avoid inappropriate exclusions? • Yes

Short Title	Title	Study Characteristics	Risk of Bias: quality assessment
			RISK Could the selection of patients have introduced bias? • High *63% of participants were stage IIIB or above.* CONCERN Is there concern that the included patients do not match the review question? • Low Were the index test results interpreted without knowledge of the results of the reference standard? • No *No blinding* If a threshold was used, was it pre-specified? • Yes RISK Could the conduct or interpretation of the index test have introduced bias? • Low Concerns regarding applicability • Low Is the reference standard likely to correctly classify the target condition? • Yes

Short Title	Title	Study Characteristics	Risk of Bias: quality assessment
			Were the reference standard results interpreted without knowledge of the results of the index test? • No *No blinding* RISK Could the reference standard, its conduct, or its interpretation have introduced bias? • High *Time is a poor reference standard because metastases could have been seeded after the initial brain imaging. The drop-out rate could have been high because there are no details of specific follow-up period(s).* CONCERN Is there concern that the target condition as defined by the reference standard does not match the review question? • Low Was there an appropriate interval between index test(s) and reference standard? • No *Time is a poor reference standard because metastases could have been seeded after the initial brain imaging.* Did all patients receive a reference standard? • Yes Did patients receive the same reference standard? • Yes

Short Title	Title	Study Characteristics	Risk of Bias: quality assessment
			Were all patients included in the analysis? • Yes RISK Could the patient flow have introduced bias? • Low Overall quality • Low
Hochsten bag 2003	MR-imaging of the brain of neurologic asymptomatic patients with large cell or adenocarcinoma of the lung. Does it influence prognosis and treatment?	**Study type** • Prospective cohort study **Study details** • Study location *The Netherlands* • Study setting *University Hospital Maastricht* • Study dates *1996 to 2000* • Duration of follow-up *For at least 6 months or until death* • Loss to follow-up *None* • Sources of funding *Not mentioned* **Inclusion criteria** • Histopathologically proven lung cancer • Staging CT chest & abdomen	**Quality assessment (cohort study)** Did the study address a clearly focused issue? • No *The study did not focus on one issue. The investigators looked at MRI brain metastases detection, the effectiveness of neurological examination to detect metastases, and the results of both these issues for people with an initial staging of I to IV.* Was the cohort recruited in an acceptable way? • Yes Was the exposure accurately measured to minimise bias? • Yes Was the outcome accurately measured to minimise bias? • No *Time is a poor reference standard because metastases could have been seeded after the initial*

Short Title	Title	Study Characteristics	Risk of Bias: quality assessment
		Exclusion criteria • Clinical evidence of remote metastases *Neurologic symptoms were not an exclusion criteria. However, the 2 people who had neurologic symptoms (and brain metastases) were initially staged as IIIB and V. Therefore, they would not affect the results of the study that we are interested in.* **Sample characteristics** • Sample size *51 people* • %female *16.5% (this percentage was calculated using the total number of people in the study, stages I to IV. However, we only analysed the results for those who had stages I to IIIA as per protocol)* • Average age *Median age (range) = 67 years (39 - 84) This percentage was calculated using the total number of people in the study, stages I to IV. However, we only analysed the results for those who had stages I to IIIA as per protocol.* **Index test / intervention (first arm of study)** • MRI brain *MRI brain was performed on a 0.5T system. First, a set of transverse spin-echo proton density and T2 weighted images was obtained with a fast spin-echo sequence (TR/TE 4500-6500/30-130 ms, FOV 230 mm, 241 x 256 matrix, NSA4, echo train length 16). T1-weighted spin-echo images were obtained (TR/TE 600/18 ms, FOV 230 mm, 10 mm slices 205 x 256 matrix, NSA4) before and after IV injection of gadolinium DTPA in a dose of 0.1 mmol/kg. One neuroradiologist interpreted the MRI brain scans.*	*brain imaging. No details were provided as to the nature of the follow-up. Follow-up MRI of the brain was performed on indication only.* Have the authors identified all important confounding factors? • Yes Have they taken account of the confounding factors in the design and/or analysis? • Yes Was the follow up of subjects complete enough? • Yes Was the follow up of subjects long enough? • Unclear *It is unclear as to what duration of follow-up is best*. Overall risk of bias • High Directness • Directly applicable **Quality assessment (diagnostic test accuracy review – QUADAS 2)** Was a consecutive or random sample of patients enrolled? • Unclear

Short Title	Title	Study Characteristics	Risk of Bias: quality assessment
		Reference standard • Followed up for at least 6 months *No details were provided as to the nature of the follow-up. Follow-up MRI of the brain was performed on indication only.* **Outcomes (study was part diagnostic, part intervention)** • Diagnostic sensitivity and specificity • Change in treatment plan: brain metastases discovered using MRI brain • Change in staging	Was a case-control design avoided? • Yes Did the study avoid inappropriate exclusions? • Yes RISK Could the selection of patients have introduced bias? • Low CONCERN Is there concern that the included patients do not match the review question? • Low Were the index test results interpreted without knowledge of the results of the reference standard? • Yes If a threshold was used, was it pre-specified? • No RISK Could the conduct or interpretation of the index test have introduced bias? • High *There was no blinding.* Concerns regarding applicability • Low

Short Title	Title	Study Characteristics	Risk of Bias: quality assessment
			Is the reference standard likely to correctly classify the target condition? • Unclear *There is no data on the ideal duration of follow-up.* Were the reference standard results interpreted without knowledge of the results of the index test? • No RISK Could the reference standard, its conduct, or its interpretation have introduced bias? • Low CONCERN Is there concern that the target condition as defined by the reference standard does not match the review question? • Low Was there an appropriate interval between index test(s) and reference standard? • No *Usually the index test and the reference standard are done with a short space of time between them. Time is a poor reference standard because metastases could have been seeded after the initial brain imaging. No details were provided as to the nature of the follow-up. Follow-up MRI of the brain was performed on indication only.* Did all patients receive a reference standard? • Yes

Short Title	Title	Study Characteristics	Risk of Bias: quality assessment
			Did patients receive the same reference standard? • Yes Were all patients included in the analysis? • Yes RISK Could the patient flow have introduced bias? • Low Overall quality • Low
Kim 2005	Screening of brain metastasis with limited magnetic resonance imaging (MRI): clinical implications of using limited brain MRI during initial staging for non-small cell lung cancer patients	**Study type** • Prospective cohort study *The second arm of this study is not relevant to our analysis because it pools data from people with stages I to IV. Therefore, for our purposes this study is effectively a prospective cohort study.* **Study details** • Study location *South Korea* • Study setting *Chungnam National University Hospital* • Study dates *Recruitment was from 2001 to 2002.* • Duration of follow-up *A minimum period of 1 year* • Loss to follow-up	**Quality assessment (cohort study)** Did the study address a clearly focused issue? • Unclear The issue being addressed does not entirely fit our protocol. However, there is data that is relevant. Was the cohort recruited in an acceptable way? • Yes Was the exposure accurately measured to minimise bias? • Yes Was the outcome accurately measured to minimise bias? • No *No reference standard*

Short Title	Title	Study Characteristics	Risk of Bias: quality assessment
		None. The 10 people who were lost to follow-up are not relevant to our analysis because the long-term follow-up data have pooled results from people who were initially staged I to IV. • Sources of funding *Not disclosed* **Inclusion criteria** • Histopathologically proven lung cancer • Staging CT chest & abdomen **Exclusion criteria** • None mentioned **Sample characteristics** • Sample size *69 people* • %female *24%. However, this is based on the initial sample size of 183 people, 114 of which are not relevant because they had an initial stage of IIIB or IV.* • Average age *Median age (range) = 67 (40 - 79). However, this is based on the initial sample size of 183 people, 114 of which are not relevant because they had an initial stage of IIIB or IV.* **Index test / intervention (first arm of study)** • MRI brain **Reference standard** • No reference standard	Have the authors identified all important confounding factors? • Yes Have they taken account of the confounding factors in the design and/or analysis? • Yes Was the follow up of subjects complete enough? • No *There was no relevant follow-up* Was the follow up of subjects long enough? • No *N/A* Overall risk of bias • High Directness • Directly applicable **Quality assessment (diagnostic test accuracy review – QUADAS 2)** Were the index test results interpreted without knowledge of the results of the reference standard? • Yes If a threshold was used, was it pre-specified? • No

Short Title	Title	Study Characteristics	Risk of Bias: quality assessment
		Although there was a follow-up period of 12 months, the investigators did not assess diagnostic accuracy. They assessed mortality. However, the mortality outcome is not relevant because they pooled data from people who had an initial stage I to IV. **Outcomes (study was part diagnostic, part intervention)** • Change in treatment plan: brain metastases discovered using MRI brain • Change in staging	RISK Could the conduct or interpretation of the index test have introduced bias? • Unclear Concerns regarding applicability • Low Is the reference standard likely to correctly classify the target condition? • Yes Were the reference standard results interpreted without knowledge of the results of the index test? • N/A RISK Could the reference standard, its conduct, or its interpretation have introduced bias? • High CONCERN Is there concern that the target condition as defined by the reference standard does not match the review question? • Low Was there an appropriate interval between index test(s) and reference standard? • No *N/A* Did all patients receive a reference standard?

Short Title	Title	Study Characteristics	Risk of Bias: quality assessment
			• No Did patients receive the same reference standard? • N/A RISK Could the patient flow have introduced bias? • High Overall quality • Low
Kormas 1992	Preoperative computed tomography of the brain in non-small cell bronchogenic carcinoma	**Study type** • Prospective cohort study **Study details** • Study location *UK* • Study setting *Hospital* • Study dates *1987 to 1989* • Duration of follow-up *12 months. However, there was no organised follow-up so the drop-out rate could have been high and false negatives could have gone undetected.* • Loss to follow-up *12 months. However, there was no organised follow-up so the drop-out rate could have been high and false negatives could have gone undetected.* • Sources of funding *Not mentioned*	**Quality assessment (cohort study)** Did the study address a clearly focused issue? • Yes Was the cohort recruited in an acceptable way? • No *There was no T staging (sizing of the primary lesion). Therefore, 0 to 13.9% of participants might not have been considered operable by UK standards.* Was the exposure accurately measured to minimise bias? • No There was no T staging Was the outcome accurately measured to minimise bias? • No *12 months. However, there was no organised follow-up so the drop-out rate could have been high and*

Short Title	Title	Study Characteristics	Risk of Bias: quality assessment
		Inclusion criteria • Histopathologically proven lung cancer **Exclusion criteria** • Evidence of metastasis in the ipsilateral chest **Sample characteristics** • Sample size *158 people* • Split between study groups *N/A* • Loss to follow-up *There was no organised follow-up so the drop-out rate could have been high and false negatives could have gone undetected.* • %female 20.3% • Average age *Mean (range): 64.1 years (40-80)* **Index test / intervention (first arm of study)** • CT brain *All CT scans were performed with 11.6 s slice time. Contrast medium (50 ml iopromide, Scherring - a non-ionic iodine based solution) was injected slowly intravenously. The total scanning time was 25 minutes. All brain computed tomograms were reported by one consultant neuroradiologist.* **Reference standard** • Outcome by 12 months	*false negatives could have gone undetected. Using time as the gold standard means that metastases could have seeded in the brain after the brain imaging, leading to erroneous false negative results.* Have the authors identified all important confounding factors? • No *Time is a poor reference standard because metastases could have been seeded after the initial brain imaging.* Have they taken account of the confounding factors in the design and/or analysis? • No *There was no organised follow-up so the drop-out rate could have been high and false negatives could have gone undetected. The presence or absence of brain metastases symptoms might affect participant behaviour with regards to whether they seek medical assistance.* Was the follow up of subjects complete enough? • No *There was no organised follow-up so the drop-out rate could have been high and false negatives could have gone undetected.* Was the follow up of subjects long enough? • Yes Overall risk of bias

Short Title	Title	Study Characteristics	Risk of Bias: quality assessment
		There was no organised follow-up so the drop-out rate could have been high and false negatives could have gone undetected. **Outcomes (study was part diagnostic, part intervention)** • Diagnostic sensitivity and specificity • Change in treatment plan: brain metastases discovered using MRI brain	• High Directness • Indirectly applicable *0% to 13.9% of participants were stage IIIB or above. 22 had incomplete T staging. Therefore, they could have been stage IIIA or stage IIIB. The reason why the results are indirectly applicable rather than partially applicable is that 7 out of the 11 brain metastases were experienced by participants in this grey area.* **Quality assessment (diagnostic test accuracy review – QUADAS 2)** Was a consecutive or random sample of patients enrolled? • No *Consecutive* Was a case-control design avoided? • Yes Did the study avoid inappropriate exclusions? • Yes RISK Could the selection of patients have introduced bias? • High *0% to 13.9% were probably inoperable (grade IIIB or above). For these participants, the probability of post-*

Short Title	Title	Study Characteristics	Risk of Bias: quality assessment
			imaging brain metastasis seeding might be higher, producing erroneous false negative results. CONCERN Is there concern that the included patients do not match the review question? • High *There was no T staging (sizing of the primary lesion). Therefore, 0 to 13.9% of participants might not have been considered operable by UK standards.* Were the index test results interpreted without knowledge of the results of the reference standard? • Yes If a threshold was used, was it pre-specified? • Yes RISK Could the conduct or interpretation of the index test have introduced bias? • Low Concerns regarding applicability • Low Is the reference standard likely to correctly classify the target condition? • No *12 month 'follow-up'. However, there was no organised follow-up so the drop-out rate could have been high and false negatives could have gone undetected.*

Short Title	Title	Study Characteristics	Risk of Bias: quality assessment
			Were the reference standard results interpreted without knowledge of the results of the index test? • No *No blinding* RISK Could the reference standard, its conduct, or its interpretation have introduced bias? • High *Time is a poor reference standard because metastases could have been seeded after the initial brain imaging.* CONCERN Is there concern that the target condition as defined by the reference standard does not match the review question? • Low Was there an appropriate interval between index test(s) and reference standard? • No *Using time as the gold standard means that metastases could have seeded in the brain after the brain imaging, leading to erroneous false negative results.* Did all patients receive a reference standard? • Unclear *No organised follow-up* Did patients receive the same reference standard? • Unclear

Short Title	Title	Study Characteristics	Risk of Bias: quality assessment
			No organised follow-up Were all patients included in the analysis? • Yes RISK Could the patient flow have introduced bias? • Low Overall quality • Low
Lee 2009	Diagnostic efficacy of PET/CT plus brain MR imaging for detection of extrathoracic metastases in patients with lung adenocarcinoma	**Study type** • Prospective cohort study **Study details** • Study location *South Korea* • Study setting *Hospital* • Study dates *2003 to 2006* • Duration of follow-up *There was no scheduled follow-up* • Loss to follow-up *n=20. These participants were excluded from the analysis* • Sources of funding *Not provided* **Inclusion criteria** • Histopathologically proven lung cancer	**Quality assessment (cohort study)** Did the study address a clearly focused issue? • Yes Was the cohort recruited in an acceptable way? • Yes Was the exposure accurately measured to minimise bias? • Yes Was the outcome accurately measured to minimise bias? • No *Time is a poor reference standard because metastases could have been seeded after the initial brain imaging.* Have the authors identified all-important confounding factors?

Short Title	Title	Study Characteristics	Risk of Bias: quality assessment
		Exclusion criteria • Unable to undergo brain MRI, for example: cardiac pacemaker, cochlear implant, intracranial aneurysm clip, known sensitivity to MRI contrast agents, presence of renal failure **Sample characteristics** • Sample size *442 people* • Split between study groups *N/A* • Loss to follow-up *Difficult to say: there was no scheduled follow-up. If participants dropped out, there was no described method to record this.* • %female *46.2% female* • Average age *Mean age (range) = 54 years (23-88)* **Index test / intervention (first arm of study)** • MRI brain *All brain MRI studies were performed by using a 3-Tesla scanner with a standard head coil. Brain MRI images were obtained in the axial, sagittal, and coronal planes by using three sequences including a T2-weighted axial turbo spin-echo pulse sequence (repetition time 3,000 ms, echo time 80 ms) with fat suppression, a fluid attenuation inversion-recovery (FLAIR) spin-echo sequence (repetition time 11,000 ms, echo time 125 ms, inversion time 2,800 ms) and a non-contrast enhanced and a contrast-enhanced T1-weighted spin-echo sequence (repetition time 500 ms, echo time 10 ms). The contrast-enhanced*	• No *They included participants who had all stages. People who were stages IIIB and IV would not have had surgery (46% of participants). This might increase the chances of metastases seeding in the brain after the initial imaging. This would increase the number who were 'false negative' erroneously. There was no scheduled follow-up. If people were to experience symptoms of a brain metastasis, this might affect their behaviour with regards as to whether they seek the assistance of the investigators or not.* Have they taken account of the confounding factors in the design and/or analysis? • No *46% of participants in this study were stage IIIB or higher. Most or all of them would not have had surgery to remove lung cancer. It is possible that for these patients there is a higher rate of erroneous false negative results because the primary cancer would remain to seed brain metastases after the initial brain imaging.* Was the follow up of subjects complete enough? • No *No scheduled follow-up* Was the follow up of subjects long enough? • Unclear *No scheduled follow-up*

Short Title	Title	Study Characteristics	Risk of Bias: quality assessment
		sequence was obtained after bolus injection of a dose of 0.2 mM/kg paramagnetic contrast agent. **Reference standard** • Outcome by 6 months *There was no follow-up schedule.* **Outcomes (study was part diagnostic, part intervention)** • Diagnostic sensitivity and specificity • Unfortunately, no other outcomes because people who were stage III were not subdivided into IIIA and IIIB	Overall risk of bias • High Directness • Indirectly applicable *35% of participants were either stage IIIA or stage IIIB. The presence of these participants might increase the false negative rate erroneously because they will not have had their lung cancer removed by surgery. In other words, they have an increased chance of seeding brain metastasis during the follow-up period. They are also more likely to receive more intensive follow-up. 28% of participants were stage IV, however, these participants' data were given separately.* **Quality assessment (diagnostic test accuracy review – QUADAS 2)** Was a consecutive or random sample of patients enrolled? • No Consecutive Was a case-control design avoided? • Yes Did the study avoid inappropriate exclusions? • No *They included participants who had all stages. People who were stages IIIB and IV would not have had surgery (46% of participants). This might increase the chances of metastases seeding in the*

Short Title	Title	Study Characteristics	Risk of Bias: quality assessment
			brain after the initial imaging. This would increase the number who were 'false negative' erroneously. There was no scheduled follow-up. If people were to experience symptoms of a brain metastasis, this might affect their behaviour with regards as to whether they seek the assistance of the investigators or not. RISK Could the selection of patients have introduced bias? • High *They included participants who had all stages. People who were stages IIIB and IV would not have had surgery (46% of participants). This might increase the chances of metastases seeding in the brain after the initial imaging. This would increase the number who were 'false negative' erroneously. There was no scheduled follow-up. If people were to experience symptoms of a brain metastasis, this might affect their behaviour with regards as to whether they seek the assistance of the investigators or not.* CONCERN Is there concern that the included patients do not match the review question? • High *28% of participants were stage IV. 35% of participants were either stage IIIA or stage IIIB. The presence of these participants might increase the false negative rate erroneously because they will not have had their lung cancer removed by surgery. In other words, they have an increased chance of seeding brain metastasis during the follow-up period.*

Short Title	Title	Study Characteristics	Risk of Bias: quality assessment
			They are also more likely to receive more intensive follow-up. Were the index test results interpreted without knowledge of the results of the reference standard? • Yes If a threshold was used, was it pre-specified? • Yes RISK Could the conduct or interpretation of the index test have introduced bias? • Low Concerns regarding applicability • Low Is the reference standard likely to correctly classify the target condition? • Unclear No scheduled follow-up Were the reference standard results interpreted without knowledge of the results of the index test? • No No blinding RISK Could the reference standard, its conduct, or its interpretation have introduced bias? • High

Short Title	Title	Study Characteristics	Risk of Bias: quality assessment
			No scheduled follow-up. Time is a poor reference standard because metastases could have been seeded after the initial brain imaging. 46% of participants in this study were stage IIIB or higher. Most or all of them would not have had surgery to remove lung cancer. It is possible that for these patients there is a higher rate of erroneous false negative results because the primary cancer would remain to seed brain metastases after the initial brain imaging. CONCERN Is there concern that the target condition as defined by the reference standard does not match the review question? • Low Was there an appropriate interval between index test(s) and reference standard? • No *Time is a poor reference standard - seeding of brain metastasis could have occurred after the initial imaging.* Did all patients receive a reference standard? • Unclear *No scheduled follow-up* Did patients receive the same reference standard? • Unclear *No scheduled follow-up* Were all patients included in the analysis? • Yes

Short Title	Title	Study Characteristics	Risk of Bias: quality assessment
			RISK Could the patient flow have introduced bias? • Low Overall quality • Low
Yohena 2004	Necessity of preoperative screening for brain metastasis in non-small cell lung cancer patients without lymph node metastasis	**Study type** • Retrospective cohort study **Study details** • Study location *Japan* • Study setting *National Kyushu Cancer Center, Fukuoka* • Study dates *1996 to 1998* • Duration of follow-up *None – this study is a snap-shot* • Loss to follow-up *N/A - this is a retrospective study that looked at patient records* • Sources of funding *Not mentioned* **Inclusion criteria** • Histopathologically proven lung cancer • Staging CT chest & abdomen **Exclusion criteria**	**Quality assessment (cohort study)** Did the study address a clearly focused issue? • Yes Was the cohort recruited in an acceptable way? • No *Retrospective study. It is hard to believe that no patient records went missing. This study had a very small number of people with brain metastases. It is possible that the records of people who had brain metastases left the cancer centre and went with them to a hospice. The date of the study is 1996 - 1998. In this period, the medical records were paper and not electronic. This is because permission was not granted for electronic records until 1999. Therefore, there was a period of 6 to 8 years between 'recruitment' and study submission for paper records (often of deceased people) to become lost.* Was the exposure accurately measured to minimise bias? • Yes Was the outcome accurately measured to minimise bias?

Short Title	Title	Study Characteristics	Risk of Bias: quality assessment
		• Those considered unresectable: T4, cN3, cM1 (except for brain metastasis) • Those considered to have some neurologic symptoms with brain metastasis **Sample characteristics** • Sample size *127 people* • %female *30.5% female. This is based on 141 people who were in the study. 14 of these people were not relevant to our protocol because 12 were T3 N2 (stage IIIB) and 2 were T4 N2 (stage IIIB).* • Average age *Mean (range) = 63 years (36 to 90). This is based on 141 people who were in the study. 14 of these people were not relevant to our protocol because 12 were T3 N2 (stage IIIB) and 2 were T4 N2 (stage IIIB).* **Index test / intervention (first arm of study)** • MRI brain **Reference standard** • No reference standard **Outcomes (study was part diagnostic, part intervention)** • Change in treatment plan: brain metastases discovered using MRI brain • Change in staging	• No *No reference standard* Have the authors identified all important confounding factors? • No *Details not provided of the MRI scanner used* Have they taken account of the confounding factors in the design and/or analysis? • Yes Was the follow up of subjects complete enough? • No Was the follow up of subjects long enough? • No No follow-up Overall risk of bias • High Directness • Directly applicable **Quality assessment (diagnostic test accuracy review – QUADAS 2)** Was a consecutive or random sample of patients enrolled? • Unclear

Short Title	Title	Study Characteristics	Risk of Bias: quality assessment
			Was a case-control design avoided? • Yes Did the study avoid inappropriate exclusions? • Unclear *Retrospective study. It is hard to believe that no patient records went missing. This study had a very small number of people with brain metastases. It is possible that the records of people who had brain metastases left the cancer centre and went with them to a hospice. The date of the study is 1996 - 1998. In this period, the medical records were paper and not electronic. This is because permission was not granted for electronic records until 1999. Therefore, there was a period of 6 to 8 years between 'recruitment' and study submission for paper records (often of deceased people) to become lost.* RISK Could the selection of patients have introduced bias? • High CONCERN Is there concern that the included patients do not match the review question? • Low Were the index test results interpreted without knowledge of the results of the reference standard? • Yes If a threshold was used, was it pre-specified?

Short Title	Title	Study Characteristics	Risk of Bias: quality assessment
			• No *There were not even details of the MRI scanner used or the time to MRI scan.* RISK Could the conduct or interpretation of the index test have introduced bias? • High Concerns regarding applicability • Low Is the reference standard likely to correctly classify the target condition? • Unclear *There is no data on how long the duration of follow-up should be.* Were the reference standard results interpreted without knowledge of the results of the index test? • No RISK Could the reference standard, its conduct, or its interpretation have introduced bias? • High CONCERN Is there concern that the target condition as defined by the reference standard does not match the review question? • Low

Short Title	Title	Study Characteristics	Risk of Bias: quality assessment
			Was there an appropriate interval between index test(s) and reference standard? • Unclear *There is no data on how long the follow-up duration should be.* Did patients receive the same reference standard? • Yes Were all patients included in the analysis? • Yes RISK Could the patient flow have introduced bias? • High Overall quality • Low
Yokoi 1999	Detection of brain metastasis in potentially operable non-small cell lung cancer: a comparison of CT and MRI	**Study type** • Prospective cohort study *This study had an MRI brain arm and a CT brain arm* **Study details** • Study location *Japan* • Study setting *Tochigi Cancer Center* • Study dates *Participants were examined with CT (CT group) between January 1989 and September 1992, and 177 participants were examined with MRI*	**Quality assessment (cohort study)** Did the study address a clearly focused issue? • Yes Was the cohort recruited in an acceptable way? • No *No explanation given as to how participants were allocated to each group.* Was the exposure accurately measured to minimise bias? • Yes

Short Title	Title	Study Characteristics	Risk of Bias: quality assessment
		(MRI group) between October 1992 and December 1995 during the 2-week period before thoracic surgery • Duration of follow-up *12 months* • Loss to follow-up *None* • Sources of funding *Not mentioned* **Inclusion criteria** • Histopathologically proven lung cancer • Staging CT chest & abdomen **Exclusion criteria** • Those considered to have some neurologic symptoms with brain metastasis • Those considered unresectable **Sample characteristics** • Sample size *332 people* • %female *MRI brain group = 34.5%; CT brain group = 29.0%* • Average age *Mean age (SD): MRI brain group = 64.8 (8.7); CT brain group = 64.2 (10.7)* **Index test / intervention (first arm of study)** • MRI brain	Was the outcome accurately measured to minimise bias? • No *For the MRI group, the final MRI was at 6 months and the clinical follow-up to 12 months. For the CT group, the final CT was at 12 months. Therefore, the outcomes were measured in a different way.* Have the authors identified all important confounding factors? • Yes Have they taken account of the confounding factors in the design and/or analysis? • Yes Was the follow up of subjects complete enough? • Yes Was the follow up of subjects long enough? • Yes Overall risk of bias • High Directness • Directly applicable **Quality assessment (diagnostic test accuracy review – QUADAS 2)**

Short Title	Title	Study Characteristics	Risk of Bias: quality assessment
		MRI images from throughout the brain were acquired with an imaging device at 1.5 T. Precontrast T1-weighted ([repetition time/echo time] 600 ms/15 ms) and T2-weighted (3,000 ms/80 ms) spin-echo axial sequences were obtained. The field of view was 20 3 20 cm. All sections were 9 mm with 1.6-mm spacing between adjacent sections, and the matrix was 256 3 256. After administering 0.2 mmol/kg gadopentetate dimeglumine, T1-weighted sequences were repeated. **Intervention 2 (second arm of study)** • CT brain *CT scans were obtained using a scanner with a 2 second scanning time. The brain was examined from the cranial base to the calvarium using 5 to 10 mm contiguous slices after IV injection of 50 mL of contrast material (iopamidol 300).* **Reference standard** • Follow-up for 12 months *Follow-up with CT and MRI was performed on people from each group who underwent complete resection of the primary tumors. Imaging was performed at the following times post-surgery: 2 months, 4 months, 6 months. In addition, the CT group underwent imaging at 12 months. The participants were scheduled for checkups every 1 to 3 months after lung resection. Furthermore, when brain metastases were suspected on examination or by the appearance of neurologic symptoms, additional scans were performed more frequently.* **Outcomes (study was part diagnostic, part intervention)** • Diagnostic sensitivity and specificity • Change in treatment plan: brain metastases discovered using MRI and CT brain	Was a consecutive or random sample of patients enrolled? • Unclear *No explanation given as to how participants were allocated to each group.* Was a case-control design avoided? • Yes Did the study avoid inappropriate exclusions? • Yes RISK Could the selection of patients have introduced bias? • Unclear *No explanation given as to how participants were allocated to each group.* CONCERN Is there concern that the included patients do not match the review question? • High *8% of participants were stage IIIB or above (non-operable)* Were the index test results interpreted without knowledge of the results of the reference standard? • Yes If a threshold was used, was it pre-specified? • Yes

Short Title	Title	Study Characteristics	Risk of Bias: quality assessment
			RISK Could the conduct or interpretation of the index test have introduced bias? • High *There was no blinding* Concerns regarding applicability • Low Is the reference standard likely to correctly classify the target condition? • Yes Were the reference standard results interpreted without knowledge of the results of the index test? • No *No blinding* RISK Could the reference standard, its conduct, or its interpretation have introduced bias? • High *Time is a bad reference standard because metastases could have been seeded after the brain imaging. For the MRI group, the final MRI was at 6 months and the clinical follow-up to 12 months. For the CT group, the final CT was at 12 months. Therefore, the outcomes were measured in a different way.* CONCERN Is there concern that the target condition as defined by the reference standard does not match the review question?

Short Title	Title	Study Characteristics	Risk of Bias: quality assessment
			• Low Was there an appropriate interval between index test(s) and reference standard? • No *Time is a bad reference standard because metastases could have been seeded after the brain imaging.* Did all patients receive a reference standard? • Yes Did patients receive the same reference standard? • No *For the MRI group, the final MRI was at 6 months. For the CT group, the final CT was at 12 months* Were all patients included in the analysis? • Yes RISK Could the patient flow have introduced bias? • Low Overall quality • Low

Appendix F – GRADE tables

Brain MRI: intervention evidence: operable people who had metastases detected by MRI brain

Quality assessment						No of patients		Effect estimate	Quality
Studies	**Design**	**Risk of bias**	**Indirectness**	**Inconsistency**	**Imprecision**	**Total**	**No found to have brain metastases using brain MRI**	**Percentage of people who were found to have brain metastases using brain MRI (95% CI)**	
Change in treatment plan: operable people who had metastases detected by MRI brain									
Earnest 1999	Prospective cohort study	Very serious[1]	Not serious	Not serious	Serious[2]	28	6	21.4% (8.3 – 41)	Very low
Hochstenbag 2003	Prospective cohort study	Very serious[1]	Not serious	Not serious	Not serious	51	5	9.8% (3.26 – 21.4)	Very low
Kim 2005	Prospective cohort study	Very serious[1]	Not serious	Not serious	Not serious	69	11	18.6% (9.69 – 30.9)	Very low
Yohena 2004	Retrospective cohort study	Very serious[1]	Not serious	Serious[3]	Not serious	127	2	1.5% (0.191 – 5.57)	Very low
Yokoi 1999	Prospective cohort study	Very serious[1]	Not serious	Not serious	Not serious	163[4]	10	6.1% (2.98 – 10.99)	Very low

1. Cohort study and has a high risk of bias. For example, time is used as the gold standard. Metastases could be seeded after the brain imaging
2. Low number of participants: between 25 and 40. Therefore, serious risk of bias
3. This outcome differs by approximately one order of magnitude compared to other studies
4. This number only includes participants who were grades I to IIIA. 14 patients inoperable according to NICE guidelines were not included (grade IIIB)

Brain MRI: intervention evidence: change in staging for people who were operable

Quality assessment									
Studies	**Design**	**Risk of bias**	**Indirectness**	**Inconsistency**	**Imprecision**	**Stage**	**Initial numbers of people according to the staging CT chest and abdomen**	**Numbers of people in these groups who had brain metastases according to MRI brain**	**Quality**
Change in staging									
Hochstenbag 2003	Prospective cohort study	Very serious[1]	Not serious	Not serious	Not serious	I	20	0	Very low
						II	12	1	
						IIIA	19	4	
Kim 2005	Prospective cohort study	Very serious[1]	Not serious	Not serious	Not serious	I	15	2	Very low
						II	16	3	
						IIIA	38	6	
Yohena 2004[2]	Retrospective cohort study	Very serious[1]	Not serious	Serious[3]	Not serious	I	76[4]	0	Very low
						II	18	1	
						IIIA	33	1	
Yokoi 1999	Prospective cohort study	Very serious[1]	Not serious	Not serious	Not serious	I	99	4[5]	Very low
						II	16	3	
						IIIA	48	3	

1. Observational study and has a high risk of bias. For example, time is used as the gold standard. Metastases could be seeded after the brain imaging
2. The original data used N & T staging. This data has been converted using the Lung Cancer Stage Grouping (8th edition)
3. This outcome differs by approximately one order of magnitude compared to other studies
4. Includes 39 people who were N0 T2. They could have been IB or IIA
5. Includes 3 people who were N0 T2. They could have been IB or IIA

Brain CT: intervention evidence: operable people who had metastases detected by CT brain

Quality assessment						No of patients		Effect estimate	Quality
Studies	Design	Risk of bias	Indirectness	Inconsistency	Imprecision	Total	No found to have brain metastases using brain CT	Percentage of people who were found to have brain metastases using brain CT (95% CI)	
Change in treatment plan: operable people who had metastases detected by MRI brain									
Yokoi 1999	Prospective cohort study	Very serious[1]	Not serious	Not serious	Not serious	143[2]	9	6.29% (2.92 – 11.6)	Very low

1. Cohort study and has a high risk of bias. For example, time is used as the gold standard. Metastases could be seeded after the brain imaging
2. This number only includes participants who were grades I to IIIA. 12 patients inoperable according to NICE guidelines were not included (grade IIIB)

Brain CT: intervention evidence: change in staging for people who were operable

Quality assessment									Quality
Studies	Design	Risk of bias	Indirectness	Inconsistency	Imprecision	Stage	Initial numbers of people according to the staging CT chest and abdomen	Numbers of people in these groups who had brain metastases according to MRI brain	
Change in staging									
Yokoi 1999	Prospective cohort study	Very serious[1]	Not serious	Not serious	Not serious	I	68	2[2]	Very low
						II	17	2	
						IIIA	58	5	

1. Observational study and has a high risk of bias. For example, time is used as the gold standard. Metastases could be seeded after the brain imaging
2. Includes 2 people who were N0 T2. They could have been IB or IIA

Diagnostic accuracy evidence: meta-analysis

No. of studies	Study design	Sample size	Sensitivity (95%CI)	Specificity (95%CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality

Brain CT									
3 (Ferrigno 1994, de Cos Escuin 2007, Kormas 1992)	Observational studies	418	74.6% (11.5 – 99.7)	99.7% (85.2 – 100.0)	Very serious[1]	Sensitivity – serious[2]; specificity – not serious	Very serious[3]	Not serious	Very low
Brain MRI									
4 (Earnest 1999, Hochstenbag 2003, de Cos Escuin 2007, Lee 2009	Observational studies	624	94.1% (68.6 – 99.9)	99.9% (91.0 – 100.0)	Very serious[1]	Sensitivity – serious[2]; specificity – not serious	Very serious[3]	Not serious	Very low

1. >33.3% of weighted data from studies at high risk of bias
2. Deviance Information Criterion was greater than 3-5 points lower for sensitivity. For specificity, the DIC was not significantly different between fixed and random effects therefore no inconsistency was observed
3. >33.3% of weighted data from studies that are indirectly relevant

Diagnostic accuracy evidence: Yokoi 1999

No. of studies	Study design	Sample size	Sensitivity (95%CI)	Specificity (95%CI)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
Brain CT									
1 (Yokoi 1999)	Observational study	155	12.5% (2.9 – 40.2%)	99.7% (96.8 – 100)	Very serious[1]	N/A	Very serious[2]	Not serious	Very low
Brain MRI									
1 (Yokoi 1999)	Observational study	177	50% (26.1 – 73.9)	99.7% (97.2 – 100)	Very serious[2]	N/A	Very serious[2]	Not serious	Very low

1. High risk of bias; note that this study was excluded from the diagnostic test accuracy meta-analysis on the grounds of clinical implausibility
2. Indirectly relevant

Appendix G – Excluded Studies

Excluded clinical studies

Short Title	Title	New column
Axelsson (2010)	An open-label, multicenter, phase 2a study to assess the feasibility of imaging metastases in late-stage cancer patients with the alpha v beta 3-selective angiogenesis imaging agent 99mTc-NC100692	All participants already had metastasis as part of the inclusion criteria. It is a study about managing brain metastasis
Hudson (2017)	Brain imaging before primary lung cancer resection: a controversial topic	No MRI/CT brain 'intervention' and no subsequent outcomes of interest that are in the protocol
Lahde (1990)	Assessing resectability of lung cancer: the role of computed tomography of the mediastinum, upper abdomen and head	Results for the presence of metastasis to the brain includes people not considered for radical treatment (e.g. CT chest & abdomen not done before recruitment to single out possible stages I to IIIA. MRI brain not done as the intervention of interest)

Li (2017)	Comparison of Gadolinium-enhanced MRI and 18FDG PET/PET-CT for the diagnosis of brain metastases in lung cancer patients: A meta-analysis of 5 prospective studies	Results for the presence of metastasis to the brain includes people not considered for radical treatment (e.g. CT chest & abdomen not done before recruitment to single out possible stages I to IIIA. MRI brain not done as the intervention of interest)
Mujoomdar (2007)	Clinical predictors of metastatic disease to the brain from non-small cell lung carcinoma: Primary tumor size, cell type, and lymph node metastases	Results for the presence of metastasis to the brain includes people not considered for radical treatment (e.g. CT chest & abdomen not done before recruitment to single out possible stages I to IIIA. MRI brain not done as the intervention of interest)
Na (2008)	A diagnostic model to detect silent brain metastases in patients with non-small cell lung cancer	Results for the presence of metastasis to the brain includes people not considered for radical treatment (e.g. CT chest & abdomen not done before recruitment to single out possible stages I to IIIA. MRI brain not done as the intervention of interest) No MRI/CT brain 'intervention' and no subsequent outcomes of interest that are in the protocol
O'Dowd (2014)	Brain metastases following radical surgical treatment of non-small cell lung cancer: is preoperative brain imaging important?	No MRI/CT brain 'intervention' and no subsequent outcomes of interest that are in the protocol
Ohno (2007)	Whole-body MR imaging vs. FDG-PET:comparison of accuracy of M-stage diagnosis for lung cancer patients	Results for the presence of metastasis to the brain includes people not considered for radical treatment (e.g. CT chest & abdomen not done before recruitment to single out possible stages I to IIIA. MRI brain not done as the intervention of interest)
Ohno (2008)	Non-small cell lung cancer: whole-body MR examination for M-stage assessment--utility for whole-body diffusion-weighted imaging compared with integrated FDG PET/CT	Results for the presence of metastasis to the brain includes people not considered for radical treatment (e.g. CT chest & abdomen not done before recruitment to single out possible stages I to IIIA. MRI brain not done as the intervention of interest)
Plathow (2008)	Positron emission tomography/computed tomography and whole-body magnetic resonance imaging in staging of advanced nonsmall cell lung cancer--initial results	Focus of the study is on whole-body MRI and does not include any outcomes of interest
Salbeck (1990)	Cerebral tumor staging in patients with	Results for the presence of metastasis to the brain includes people not considered for radical

	bronchial carcinoma by computed tomography	treatment (e.g. CT chest & abdomen not done before recruitment to single out possible stages I to IIIA. MRI brain not done as the intervention of interest)
Salvatierra (1990)	Extrathoracic staging of bronchogenic carcinoma	Results for the presence of metastasis to the brain includes people not considered for radical treatment (e.g. CT chest & abdomen not done before recruitment to single out possible stages I to IIIA. MRI brain not done as the intervention of interest)
Seute (2008)	Detection of brain metastases from small cell lung cancer: consequences of changing imaging techniques (CT versus MRI)	Results for the presence of metastasis to the brain includes people not considered for radical treatment (e.g. CT chest & abdomen not done before recruitment to single out possible stages I to IIIA. MRI brain not done as the intervention of interest)
Suzuki (2004)	Magnetic resonance imaging and computed tomography in the diagnoses of brain metastases of lung cancer	Results for the presence of metastasis to the brain includes people not considered for radical treatment (e.g. CT chest & abdomen not done before recruitment to single out possible stages I to IIIA. MRI brain not done as the intervention of interest)
van de Pol (1996)	MRI in detection of brain metastases at initial staging of small-cell lung cancer	Results for the presence of metastasis to the brain includes people not considered for radical treatment (e.g. CT chest & abdomen not done before recruitment to single out possible stages I to IIIA. MRI brain not done as the intervention of interest)

Excluded economic studies

Paper	Primary reason for exclusion
Colice, G.L., Birkmeyer, J.D., Black, W.C., Littenberg, B. and Silvestri, G., 1995. Cost-effectiveness of head CT in patients with lung cancer without clinical evidence of metastases. Chest, 108(5), pp.1264-1271.	Study conducted in a non-UK setting.
Wernicke, A. Gabriella, Menachem Z. Yondorf, Bhupesh Parashar, Dattatreyudu Nori, KS Clifford Chao, John A. Boockvar, Susan Pannullo, Philip Stieg, and Theodore H. Schwartz. "The cost-effectiveness of surgical resection and cesium-131 intraoperative brachytherapy versus surgical resection and stereotactic radiosurgery in the treatment of metastatic brain tumors." Journal of neuro-oncology 127, no. 1 (2016): 145-153	Study did not include suitable comparators.

Appendix H – References

Clinical Studies - Included

Earnest F, Ryu J H, Miller G M, Luetmer P H, Forstrom L A, Burnett O L, Rowland C M, Swensen S J, and Midthun D E (1999) Suspected non-small cell lung cancer: incidence of occult brain and skeletal metastases and effectiveness of imaging for detection--pilot study. Radiology 211(1), 137-45

Hochstenbag M M, Twijnstra A, Hofman P, Wouters E F, ten Velde, and G P (2003) MR-imaging of the brain of neurologic asymptomatic patients with large cell or adenocarcinoma of the lung. Does it influence prognosis and treatment?. Lung Cancer 42(2), 189-93

Kim S Y, Kim J S, Park H S, Cho M J, Kim J O, Kim J W, Song C J, Lim S P, and Jung S S (2005) Screening of brain metastasis with limited magnetic resonance imaging (MRI): clinical implications of using limited brain MRI during initial staging for non-small cell lung cancer patients. Journal of Korean Medical Science 20(1), 121-6

Yohena T, Yoshino I, Kitajima M, Uehara T, Kanematsu T, Teruya T, Ikeda J, and Ichinose Y (2004) Necessity of preoperative screening for brain metastasis in non-small cell lung cancer patients without lymph node metastasis. Annals of Thoracic & Cardiovascular Surgery 10(6), 347-9

Clinical studies – Excluded

Axelsson R, Bach-Gansmo T, Castell-Conesa J, McParland B J, and Study Group (2010) An open-label, multicenter, phase 2a study to assess the feasibility of imaging metastases in late-stage cancer patients with the alpha v beta 3-selective angiogenesis imaging agent 99mTc-NC100692. Acta Radiologica 51(1), 40-6

de Cos Escuín, J S, Menna D M, González M A, Quirantes J Z, Vicente C D, and Calvo M C (2007) Silent brain metastasis in the initial staging of lung cancer: evaluation by computed tomography and magnetic resonance imaging. Arch Bronconeumol 43, 386-91

Ferrigno D, and Buccheri G (1994) Cranial computed tomography as a part of the initial staging procedures for patients with non-small-cell lung cancer. Chest 106(4), 1025-9

Hudson Z, Internullo E, Edey A, Laurence I, Bianchi D, and Addeo A (2017) Brain imaging before primary lung cancer resection: a controversial topic. Ecancermedicalscience 11, 749

Kormas P, Bradshaw J R, and Jeyasingham K (1992) Preoperative computed tomography of the brain in non-small cell bronchogenic carcinoma. Thorax 47(2), 106-8

Lahde S, Paivansalo M, Rainio P, Merikanto J, and Karkola P (1990) Assessing resectability of lung cancer: the role of computed tomography of the mediastinum, upper abdomen and head. European Journal of Radiology 10(1), 48-55

Lee H Y, Lee K S, Kim B T, Cho Y S, Lee E J, Yi C A, Chung M J, Kim T S, Kwon O J, and Kim H (2009) Diagnostic efficacy of PET/CT plus brain MR imaging for detection of extrathoracic metastases in patients with lung adenocarcinoma. Journal of Korean Medical Science 24(6), 1132-8

Li Y, Jin G, and Su D (2017) Comparison of Gadolinium-enhanced MRI and 18FDG PET/PET-CT for the diagnosis of brain metastases in lung cancer patients: A meta-analysis of 5 prospective studies. Oncotarget 8(22), 35743-35749

Mujoomdar A, Austin J H. M, Malhotra R, Powell C A, Pearson G D. N, Shiau M C, and Raftopoulos H (2007) Clinical predictors of metastatic disease to the brain from non-small

cell lung carcinoma: Primary tumor size, cell type, and lymph node metastases. Radiology 242(3), 882-888

Na II, Lee T H, Choe D H, Cheon G J, Kim C H, Koh J S, Baek H, Ryoo B Y, Yang S H, and Lee J C (2008) A diagnostic model to detect silent brain metastases in patients with non-small cell lung cancer. European Journal of Cancer 44(16), 2411-7

O'Dowd E L, Kumaran M, Anwar S, Palomo B, and Baldwin D R (2014) Brain metastases following radical surgical treatment of non-small cell lung cancer: is preoperative brain imaging important?. Lung Cancer 86(2), 185-9

Ohno Y (2007) Whole-body MR imaging vs. FDG-PET:comparison of accuracy of M-stage diagnosis for lung cancer patients. J Magn Reson Imaging 26(3), 498-509

Ohno Y (2008) Non-small cell lung cancer: whole-body MR examination for M-stage assessment--utility for whole-body diffusion-weighted imaging compared with integrated FDG PET/CT. Radiology 248(2), 643-54

Plathow C (2008) Positron emission tomography/computed tomography and whole-body magnetic resonance imaging in staging of advanced nonsmall cell lung cancer--initial results. Invest Radiol 43(5), 290-7

Salbeck R, Grau H C, and Artmann H (1990) Cerebral tumor staging in patients with bronchial carcinoma by computed tomography. Cancer 66(9), 2007-11

Salvatierra A, Baamonde C, Llamas J M, Cruz F, and Lopez-Pujol J (1990) Extrathoracic staging of bronchogenic carcinoma. Chest 97(5), 1052-8

Seute T, Leffers P, ten Velde, G P, and Twijnstra A (2008) Detection of brain metastases from small cell lung cancer: consequences of changing imaging techniques (CT versus MRI). Cancer 112(8), 1827-34

Suzuki K, Yamamoto M, Hasegawa Y, Ando M, Shima K, Sako C, Ito G, and Shimokata K (2004) Magnetic resonance imaging and computed tomography in the diagnoses of brain metastases of lung cancer. Lung Cancer 46(3), 357-60

van de Pol , M , van Oosterhout , A G, Wilmink J T, ten Velde, G P, and Twijnstra A (1996) MRI in detection of brain metastases at initial staging of small-cell lung cancer. Neuroradiology 38(3), 207-10

Yokoi K, Kamiya N, Matsuguma H, Machida S, Hirose T, Mori K, and Tominaga K (1999) Detection of brain metastasis in potentially operable non-small cell lung cancer: a comparison of CT and MRI. Chest 115(3), 714-9

Health Economic studies – Included

None

Health Economic studies – Excluded

Colice, G.L., Birkmeyer, J.D., Black, W.C., Littenberg, B. and Silvestri, G., 1995. Cost-effectiveness of head CT in patients with lung cancer without clinical evidence of metastases. *Chest*, *108*(5), pp.1264-1271.

Wernicke, A. Gabriella, Menachem Z. Yondorf, Bhupesh Parashar, Dattatreyudu Nori, KS Clifford Chao, John A. Boockvar, Susan Pannullo, Philip Stieg, and Theodore H. Schwartz. "The cost-effectiveness of surgical resection and cesium-131 intraoperative brachytherapy versus surgical

resection and stereotactic radiosurgery in the treatment of metastatic brain tumors." *Journal of neuro-oncology* 127, no. 1 (2016): 145-153.

Appendix I – Cost-utility analysis

Background

Brain metastases (BM) are a frequent complication from non-small cell lung cancer (NSCLC) but routine imaging of the brain is not undertaken, especially in early stage disease. The 2011 guideline included a recommendation to "Consider MRI or CT of the head in patients selected for treatment with curative intent, especially in stage III disease" but it is not known how widely this guidance is implemented in UK practice or whether practice differs by cancer stage. Detecting BM prior to treatment with curative intent is valuable as it may alter the treatment plan. For example, patients initially indicated for surgery may be switched to less invasive treatment as the chance for cure is greatly reduced if they are found to have BM. Early detection of BM may also lead to better outcomes for patients in that they may be able to receive BM-specific treatment that will better their prognosis.

The prevalence of BM is thought to be relatively low in patients with early stage NSCLC and, given that CT and MRI have limited availability, the committee were interested in examining the cost-effectiveness of routine imaging separately in patients with stage I, II and III disease. An important motivator for the inclusion of this review question in the guideline update was the publication of the O'dowd 2014 paper[b], which tried to estimate the prevalence of BM in the population of interest.

Methods

Population, interventions/comparators and outcomes

The populations in the model are patients with stage I, II and III NSCLC who are otherwise selected for treatment with curative intent; either surgery or radical radiotherapy to the lung. These patients have already received the standard lung cancer staging investigations of chest CT, whole body non-contrast-enhanced PET-CT and any necessary biopsy procedures. The cancer stage is expected to be correct in all respects except for the potential for occult BM. Patients in the model are either negative for BM, are positive with 1-3 BM or are positive with 4+ BM. The distinction is clinically important in that patients with 1-3 BM often receive radical brain treatment and then may go on to receive radical treatment to their lung whereas patients with 4+ BM receive treatment that is systemic and palliative in nature.

The strategies examined in the model were No Imaging (i.e. straight to radical treatment), CT of the brain followed by MRI if positive and MRI of the brain alone. Outcomes were measured in quality adjusted life years (QALYs).

[b] O'Dowd et al (2014) Brain metastases following radical surgical treatment of non-small cell lung cancer: is preoperative brain imaging important? Lung Cancer. 2014 Nov;86(2):185-9

Model Structure

This section is intended to give a structural overview of the model and its underpinning assumptions. Derivation of parameters is discussed in the Model Parameters section.

Short Term Model

The model begins with a series of decision trees which determine the results of the diagnostic tests undertaken on 1,000 theoretical patients. Following this, patients have the potential to be either True Negative (TN), True Positive (1-3), True Positive (4+), False Positive (1-3) or False Negative (4+). There are no TPs or FPs in the No Imaging strategy as no test has taken place. In Figure 1, p is the prevalence of BM, pr(1-3) is the proportion of patients with BM that have 1-3 BM, pr(4+) is the proportion of patients with BM that have 4+ BM, seMRI and spMRI are the sensitivity and specificity of MRI. Sensitivity is expected to be higher for patients with 4+ BM.

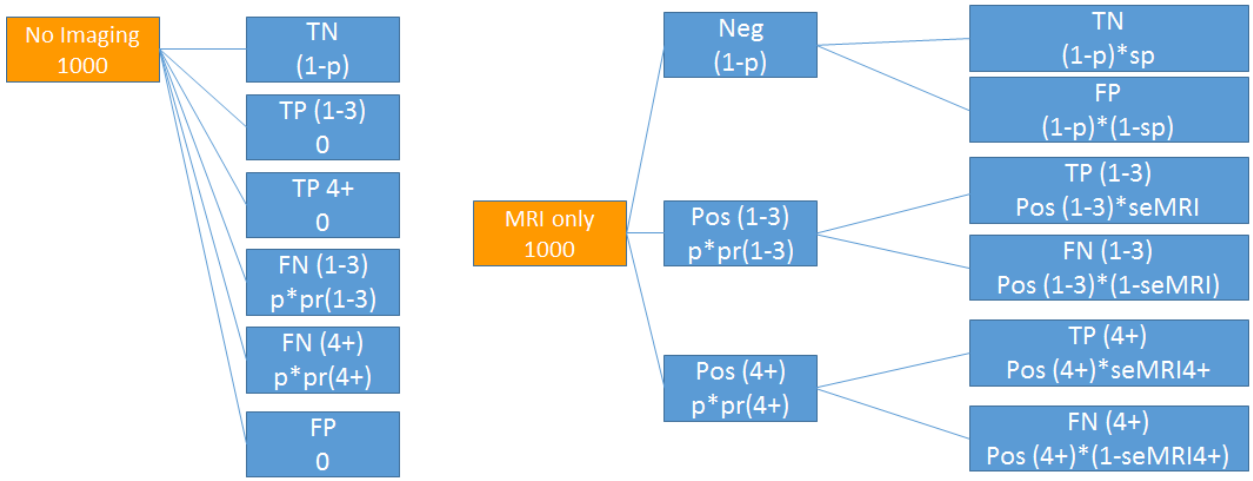


Figure 1: Diagnostic Decision Trees (No Imaging and MRI only Strategies)

In Figure 2, p is the prevalence of BM, pr(1-3) is the proportion of patients with BM that have 1-3 BM, pr(4+) is the proportion of patients with BM that have 4+ BM, sect, seMRI, spCT and spMRI are the sensitivity and specificity of CT and MRI. Sensitivity is expected to be higher for patients with 4+ BM. Patients who are identified as positive (4+) do not receive a confirmatory MRI in the base case analysis.

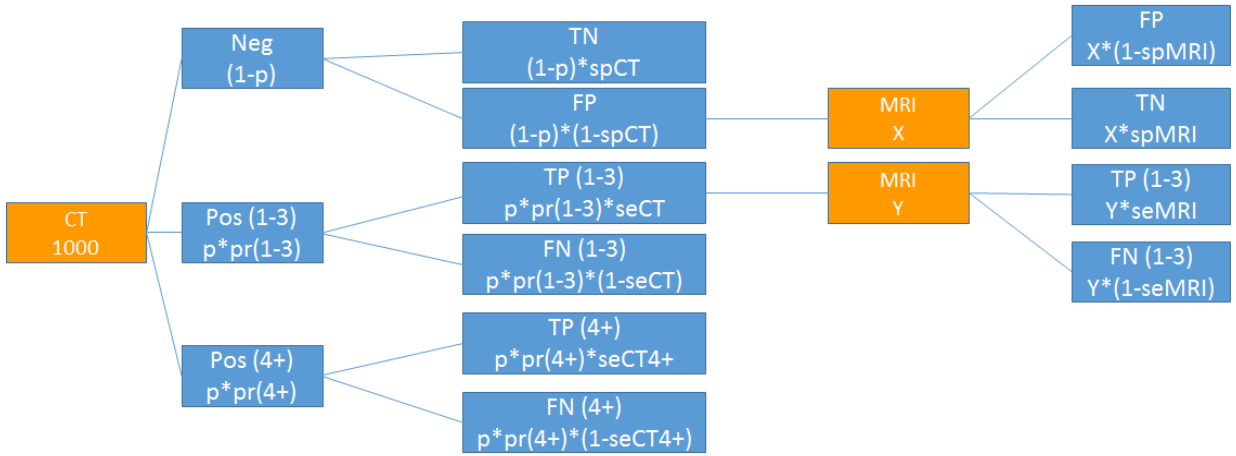


Figure 2: Diagnostic Decision Tree for CT-MRI Strategy

Following initial imaging, those patients who are found to be negative receive the treatment with curative intent that they had initially been indicated for (comprising various types of surgery and radical radiotherapy). Many of the patients who are found to be positive (1-3) receive radical treatment for both their brain metastases and on their lung. Patients who are found to be positive (4+) are assumed to receive treatments that are systemic or palliative in nature rather than radical. The exact breakdown of these treatments is discussed in the parameters section of this report. An important assumption of this analysis is that specific treatments do not affect patients' prognoses. The reason for this assumption is that both the patient group under study and their treatment options are very heterogeneous so the model would have quickly become unmanageably complicated and would have required a large number of parameters for which data do not exist. We therefore chose to model broad groups of patients for whom robust data do exist based on the outcomes of the diagnostic tests.

We assume that the sensitivity and specificity of MRI, when used in the MRI alone strategy is the same when used on the patient population who have been confirmed as positive with CT scanning. The committee indicated this assumption was reasonable. Another important assumption of the model is that the testing strategies do not generate any genuine False Positives. This is because the specificity of MRI for detecting brain metastases was found to be ~100% in the clinical review. The committee stated that they believe this to be true; while evidence from the clinical review showed MRI scans identifying phenomena that mimicked lesions such as "flow related enhancements" and may detect differential diagnoses such as primary brain tumours or infections, the committee were of the view that MRI would not falsely detect brain metastases and that a patient would not be managed as if they had brain metastases when they did not. While there

might be the odd highly unusual contradiction to this assumption, for the purposes of the model it was reasonable to assume there were no False Positives.

If there are no False Positives then there is no value in modelling True Negatives as the numbers will not differ by strategy. Therefore all True Negative patients exit the model after initial imaging. Included in the True Negative patients exiting the model are those few patients with differential diagnoses such as primary brain tumours and infections as it is assumed they will be managed in a cost-effective way elsewhere for those conditions, in addition to receiving appropriate treatment for their NSCLC. It was thought this cohort are small enough that the potential gain in net monetary benefit from incidentally identifying them via imaging was assumed not to affect the conclusions of the model.

The diagnostic decision trees and the initial treatments that patients receive are assumed, for the purposes of the model, to occur instantaneously. That is, there are no negative effects from delay due to imaging and all patients are assumed to receive some initial treatment before any deaths or progressions occur. Addressing this limitation would have required a number of evidence-free assumptions about the effects of delay that would have likely only had a minor effect on the results.

Long Term Model

At the end of the diagnostic decision tree there are four broad patient groups to model the outcomes for; True Positives (1-3), True Positives (4+), False Negatives (1-3) and False Negative (4+), all of whom have BM. A Partitioned Survival Analysis[c] (PartSA) model was chosen as it is the most common structure for modelling advanced cancers and due to the availability of relevant data to calculate the model's parameters. A PartSA model makes use of overall (OS) and progression free survival (PFS) curves to partition patients into three mutually exclusive states at any given point in time; 'dead', 'alive and progression free', 'alive and progressed'. At each time point the proportion of patients in the dead state is given by one minus the overall survival curve, the proportion of patients alive and progression free is equal to the progression free survival curve and therefore the number alive and progressed is equal to one minus the sum of the other two groups.

The model is a state membership rather than a state transition model so some assumptions are needed to model transition related events. It would not possible for any patient to transition from the progressed to the progression free state but it would be possible for a patient to transition from the progression free state to either the progressed state and for patients in either state to transition to the dead state. The number of transitions assumed to occur to the dead state from cycle to cycle is therefore equal to the difference in the dead state membership and the number of transitions from the progression free to the progressed state is equal to the difference in the progression free state minus the number of first events that are deaths (these data need to be obtained from trials). Both types of transition events incur important one off costs in NSCLC patients so it was necessary to characterise their occurrences explicitly in this way. Figure 3 shows how the OS and PFS curves dictate the proportions of patients in each state in a typical PartSA model.

[c] NICE DSU TSD 19: Partitioned survival analysis for decision modelling in health care: a critical review (2007)

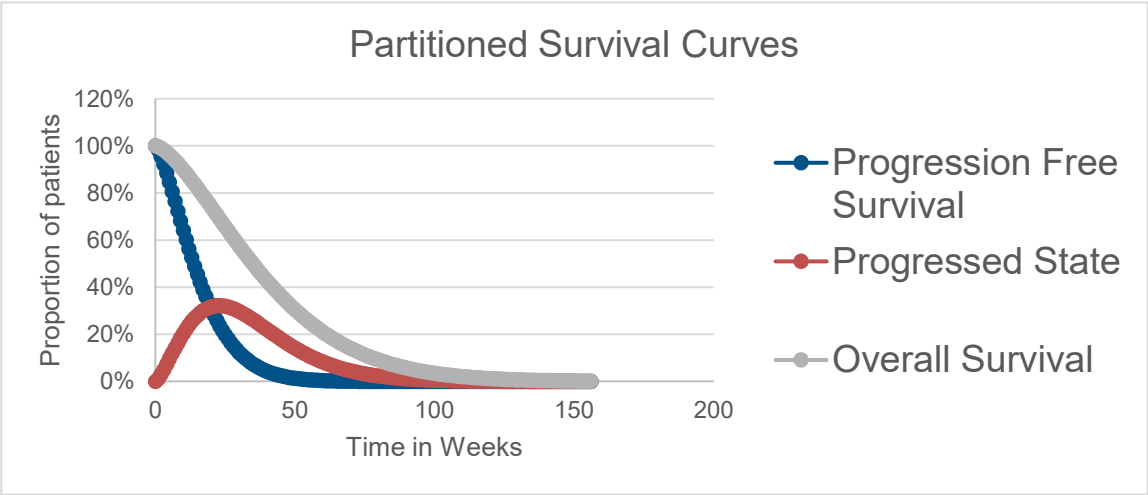


Figure 3: Typical Partitioned Survival Analysis Model State Membership

Overall Survival

For this specific model True Positives (1-3) were assumed to proceed along an OS curve that was obtained from trials of relevant patients. The OS of patients who were True Positive (4+) was calculated by applying a hazard ratio (HR) relevant to the proportional hazard between these two groups. As is often the case in diagnostic models that try to capture the outcomes for patients with an incorrect diagnosis, some strong structural assumptions were needed for the False Negative cohorts. Patients who were False Negative (1-3) were assumed to begin with a HR of 1 versus the TP (1-3) group and were then assumed to gradually progress to having an equal HR to the TP (4+) group over the average time to intracranial progression in a trial of patients with BM multiplied by 2 (it was assumed that the vast majority of patients would have developed 4+ BM by this time). Evidence on the natural history of BM from the O'Dowd paper as well as the trials used to inform parameters in this model lent credence to the assumption that BM grow and proliferate over a relatively short time period. The committee confirmed that this assumption was reasonable, given their clinical experience of managing these patients. The overall survival curve for patients who were FN4+ was calculated by applying an initial hazard ratio representative of Whole Brain Radio Therapy (WBRT) treatment to the overall survival curve for patients who were TP4+. This parameter was taken from an RCT for use of WBRT versus best supportive care in patients with a good performance status and brain metastases from NSCLC (Mulvenna et al. 2016). Because the TP4+ patients were treated with WBRT and the FN4+ patients were not, we considered this a reasonable approximation. The hazard ratio declined uniformly, cycle by cycle (to represent patients gradually presenting symptomatically) and

became equal to one at two times the average time to presentation, by which time all patients who were likely to progress intracranially were assumed to have progressed.

Progression-Free Survival

Progressions were defined as either intracranial or extracranial or both together and could occur at initial or distant sites or both together. Data on the PFS curve was obtained from a trial of relevant patients with 1-3 BM (Kocher et al. 2011). This PFS curve was used directly for the cohort who were TP (1-3) but a series of assumptions needed to be made to translate it to the other groups. The PFS curve for TP (1-3) was divided by the OS curve for TP (1-3) to give a proportion alive and progression free at each time point, this was multiplied by the OS curve for TP (4+) to give the PFS curve for TP (4+). It was assumed that the difference in survival between patients that were TP (1-3) and TP (4+) was directly attributable to BM. The committee thought this a reasonable assumption as the multivariable regression that had provided the relevant hazard ratio had controlled for other patient level factors. This assumption then extends to the difference in OS for the FN (1-3) group. To try to approximate this relationship, the model accelerated the PFS curve by an acceleration factor that would ensure the absolute difference in the area under the FN (1-3) and TP (1-3) PFS curves from time point 0 to 42 weeks (as discussed earlier, this was the point at which all intracranial progressions in the FN group were assumed to have occurred) equal to the absolute difference in the area under the corresponding OS curves. This assumption was tested in sensitivity analysis. We applied the same logic for the (4+) population as the (1-3) population for PFS, accelerating the PFS curve for FN (4+) to a value that ensured the absolute difference in the area under the curve between the FN (4+) and TP (4+) population was equal to the difference in their corresponding overall survival curves at 42 weeks. The combination of acceleration factors and the multiplicative approach to PFS curves has the advantage of preserving the relationship of PFS and OS in the different patient groups and in sensitivity analyses but the disadvantage that there will be a very small amount of “double-counting” progressions following 42 weeks. Because the PFS curve will still be multiplied by a lower OS curve but the internal logic of the model is that the FN patients who are going to progress are assumed to progress by this point and that all differences in OS are attributable to PFS, a lower PFS curve beyond 42 weeks is perhaps inconsistent. It can be seen in Figure 4 that the effect of this is a very minimal, however, and might reflect a clinically reasonable ‘tail’ of late progression.

Figure 4 shows a diagram of the structure of the partitioned survival analysis component of the economic model.

The average age at the start of the model was 60 (the average age in relevant BM trials), the model was run on a weekly cycle length for 10 years in the base case. While a few patients were left alive at the end of the time horizon, the committee were mindful that every patient within the model has NSCLC and BM and found it highly unlikely that anyone would survive beyond this time point. Due to small patient numbers, this issue was not expected to meaningfully affect the conclusions of the model, however.

Patients existing in the progression free and progressed states accrued QALYs as a multiple of relevant utility values and time in state. They also accrued routine NSCLC management costs for existing in both states. Progression and death events both accrued one-off event costs, which are discussed in more detail in the Model Parameters section.

Both costs and QALYs were discounted at 3.5% and a half cycle correction was applied.

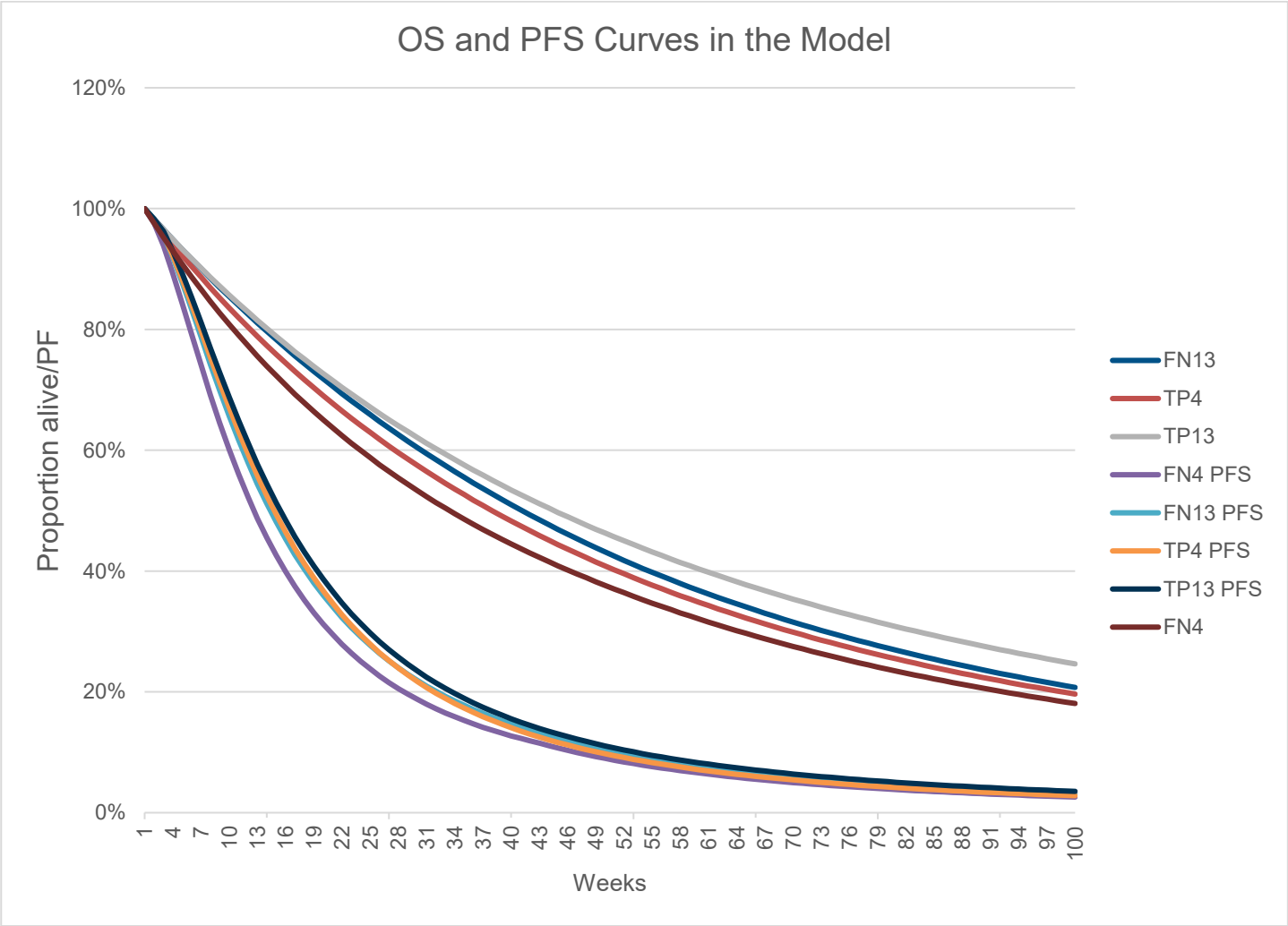


Figure 4: Structure of the Long Term Model

Model Parameters

Prevalence of BM

As stated in the section detailing the model structure, the prevalence of BM in the three populations of interest was obtained from a paper by O'Dowd et al 2014. This was a retrospective study of 646 NSCLC patients undergoing treatment with curative intent at a UK hospital so was seen by the committee as the most relevant source of data for this parameter. The analysis included 41 patients who had been identified as having BM in a maximum follow up period of 2 years. The size of the BM and a tumour doubling time of 58.48 days[d] were used to estimate the proportion of patients who had BM at the time of their radical treatment. The paper estimated that 71% of these metastases were above 5mm in diameter and 83% were above 2mm. The committee felt that the 2mm cut-off was the more relevant for modern MRI scanners but the 5mm cut-off was used in sensitivity analysis. The prevalence values were multiplied by the proportion detectable to calculate the proportion of detectable BM in the model.

Table 3: Prevalence of BM

Parameter	Value	Lower CI	Upper CI	Source
Stage I - Prevalence of BM	4.6%	2.7%	7.1%	O'Dowd 2014
Stage II - Prevalence of BM	9.5%	5.3%	14.8%	O'Dowd 2014
Stage IIIA - Prevalence of BM	9.3%	4.6%	15.5%	O'Dowd 2014
Proportion detectable (2mm)	83%			O'Dowd 2014
Proportion detectable (5mm)	71%			O'Dowd 2014

Based on the natural history of NSCLC, one would expect the prevalence of BM to be higher in stage IIIA than in stage II. The equivalence observed in this data could be due to the patients having received a staging PET-CT, which could have detected the larger and more obvious BM and therefore ruled them out of receiving radical treatment. The patients in this study occupy the same point in the care pathway as the patients in this decision problem so the committee thought the data were directly relevant but recognised that in centres that use contrast enhanced PET-CT at initial staging, the prevalence of BM might be lower.

Diagnostic Test Accuracy

Sensitivity (Se) is the probability that a diagnostic test will correctly identify a positive patient as positive. Specificity (Sp) is the probability that a diagnostic test will correctly identify a negative patient as negative. In order to determine these parameters, we used studies reporting the relevant data that had been identified as part of the clinical sift for this question. The relevant data are in Table 4.

[d] Yoo H, Nam B-H, Yang H-S, Shin SH, Lee JS, Lee SH. Growth rates of metastatic brain tumors in non-small cell lung cancer. Cancer 2008;113(5):1043–7.

Table 4: Diagnostic Test Accuracy of CT and MRI

Study	Modality	Prevalence	% IIIB and above	Negatives	Positives	N	Sensitivity	Specificity
Ferrigno 1994	CT	14%	63%	159	25	184	92%	99%
de Cos Escuin 2007	CT	8%	69%	70	6	76	67%	100%
Kormas 1992	CT	6%	-	149	9	158	44%	99%
Yokoi 1999	CT	7%	8%	144	11	155	9%	100%
Earnest 1999	MRI	21%	0%	23	6	29	100%	100%
Hochstenbag 2003	MRI	9%	0%	51	5	56	100%	100%
de Cos Escuin 2007	MRI	11%	55%	86	11	97	91%	100%
Lee 2009	MRI	10%	46%	399	43	442	86%	98%
Yokoi 1999	MRI	7%	8%	165	12	177	50%	100%

There are a number of limitations to these studies; several were old and therefore used out of date equipment, there was a relatively significant prevalence of patients with stage IIIB NSCLC and above in the studies (although the committee assessed this limitation as minor as regards the accuracy of the tests), there were a small number of positive patients on which to base the sensitivity calculations and the method for determining sensitivity was of varying quality. Nevertheless, these were the only empirical data available and the committee were content to use them in the base case analysis. For this base case, they decided to exclude the data from Yokoi 1999 as the sensitivity values looked implausibly low at 9% for CT and 50% for MRI.

We performed independent meta-analyses for Se and Sp for both MRI and CT using WinBUGS. We attempted to fit bivariate models (i.e. where Se and Sp were correlated) but did not have enough studies for the MCMC algorithm to be stable. The WinBUGS code can be found in Appendix L and the results are in Table 5.

Table 5: Results of DTA Meta-Analyses

		MRI			DIC	CT			DIC
Incl Yokoi 1999		LowCI	**Estimate**	HighCI		LowCI	**Estimate**	HighCI	
Random effects	Sensitivity	40.2%	92.0%	100.0%	17.06	1.8%	55.0%	98.7%	17.374
Fixed Effects		74.0%	83.4%	90.6%	22.25	49.2%	62.9%	75.4%	38.088
Random effects	Specificity	98.3%	100.0%	100.0%	7.191	96.7%	99.9%	100.0%	9.126
Fixed Effects		98.2%	99.1%	99.6%	14.21	98.9%	99.7%	100.0%	8.141
Excl Yokoi 1999									
Random effects	Sensitivity	68.6%	94.1%	99.9%	11.19	11.5%	74.6%	99.7%	13.665
Fixed Effects		80.7%	89.6%	95.5%	10.41	63.5%	78.0%	88.9%	18.198
Random effects	Specificity	91.0%	99.9%	100.0%	7.531	85.2%	99.7%	100.0%	8.143
Fixed Effects		50.6%	95.0%	100.0%	10.56	98.5%	99.6%	99.9%	6.844

The committee chose to prefer random effects models for Se and Sp for both CT and MRI, which reflected a combination of the heterogeneity of the studies and the DIC statistics. This gave Se values of 74.6% for CT and 94.1% for MRI and Sp values of 99.7% for CT and 99.9% for MRI.

The committee examined the data on False Positives in the underpinning studies and decided that they were not relevant to current practice, particularly for MRI. This was because the source of False Positives in the Lee 2009 study was listed as 'flow related enhancements', which are thought to no longer be a factor. The committee agreed that in their experience there would be no genuine False Positives (i.e. those that would lead to someone being treated for BM when they did not, in fact, have BM) following an MRI scan. As discussed earlier, differential diagnosis, while a consequence of imaging were not expected to affect the conclusions of the model due to small numbers. A specificity value of 100% (rather than 99.9%) was therefore used in the model and because there were no False Positives in any of the strategies, long term outcomes for False Positives and True Negatives were not modelled. This value was necessarily fixed at 100% in the probabilistic sensitivity analysis.

In the base case, the Se of both CT and MRI for detecting people with 4+ BM was fixed at 100% on the advice of the committee. While this assumption was relaxed in sensitivity analysis for CT, the committee thought it highly implausible that MRI would not detect someone with 4+ BM of above 2mm in diameter.

Number of BM

As discussed in the model structure section, the committee indicated that the number of brain metastases identified could significantly alter subsequent treatment decisions. They specified two broad patient groups of interest, those who had 1-3 BM and those who had 4+ BM. The committee's *a priori* assumption was that 90% of positive patients would have 1-3 BM. We also identified data in a relevant population[e] showing the proportion to be 74% (CI 55% - 89%). These data, while quite uncertain, are very important in the model as the initial treatments received by patients with 1-3 BM are far higher in cost than those received by the patients with 4+ BM. Therefore, the higher we believe the proportion of patients with 4+ BM to be, the more cost-effective imaging will be. In the base case, we used the 74% value for the number of positive patients would have 1-3 BM, examining the effects of the 90% value in sensitivity analysis.

Survival Curve Parameter Estimation Method

All survival curve parameters used in the model were obtained from studies using the algorithm from Guyot 2012[f]. The algorithm makes use of Kaplan-Meier (KM) curves that are digitised using graph digitisation software (Enguage[g] was used for this purpose) and the numbers at risk (often published beneath KM curves in studies) at various time points to estimate synthetic individual patient survival and censorship data. The synthetic individual patient data is then amenable to survival analysis and statistics such as hazard ratios and parametric survival curve parameters may be obtained in the normal way. STATA[h] was used for this purpose. This method has been extensively validated, with survival analysis statistics generated using synthetic data very closely mirroring those produced using the relevant real trial data in a large number of examples (see also Guyot 2012).

[e] Yokoi et al *Chest*. 1999 Mar;115(3):714-9.
[f] Guyot et a (2012) Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. BMC Medical Research Methodology

[g] http://digitizer.sourceforge.net/

[h] https://www.stata.com/

Overall Survival Curves

No fully direct data were identified that would have enabled us to estimate survival curves for the populations of interest within the model. Instead a number of partially applicable studies were discussed with the committee:-

- Kocher 2011[i], an RCT in a European setting that investigated Whole Brain Radiotherapy (WBRT) + Radical Treatment versus Radical Treatment alone in patients with 1-3 brain metastases (only 53% of whom had NSCLC). N=359
- Brown 2016[j], an RCT in a US setting that investigated WBRT + Stereotactic Radiosurgery (SRS) versus SRS alone for people with 1-3 brain metastases (only 69% of whom had 'lung' cancer). N=213
- Sperduto 2016[k], a retrospective study in a US setting that estimated prognostic indicators for the survival of people with NSCLC and brain metastases. N=2,186
- The IASLC Lung Cancer Staging Project 2015, a retrospective study in a European setting that underpinned the TNM8 NSCLC staging criteria. N=1,059

The most relevant data from the Sperduto study were survival curves relating to the group with a GPA 2.5-3 (age under 70, good Karnofsky Performance Status, absent of extracranial metastases 1-4 BM and EGFR/ALK status unknown). The committee discussed all the relevant survival curves and the strengths and limitations of the studies. They concluded that the IASLC TNM8 data only included sparse data on people with BM so should be excluded from the analysis but were unable to decide which of the Kocher, Brown and Sperduto studies was the most relevant to the patient group who were True Positive (1-3). For Kocher and Brown, the study arms that did not receive WBRT were used as this is not standard treatment for people with 1-3 BM. The committee noted that the Kocher and Brown studies had been used in the economic model conducted for NICE's Guideline on Brain tumours and brain metastases[l] and that the median and interquartile range values for all three curves were similar and clinically plausible. They therefore requested that the OS curve in the model should be based on a meta-analysis of all three.

[i] Kocher et al (2011) Adjuvant Whole-Brain Radiotherapy Versus Observation After Radiosurgery or Surgical Resection of One to Three Cerebral Metastases: Results of the EORTC 22952-26001 Study. Journal of Clinical Oncology
[j] Brown et al (2016) Effect of Radiosurgery Alone vs Radiosurgery With Whole Brain Radiation Therapy on Cognitive Function in Patients With 1 to 3 Brain Metastases. JAMA
[k] Sperduto et al (2016) Estimating Survival in Patients With Lung Cancer and Brain Metastases. JAMA Oncology
[l] The National Institute for Health and Care Excellence (2018). Brain tumours (primary) and brain metastases in adults

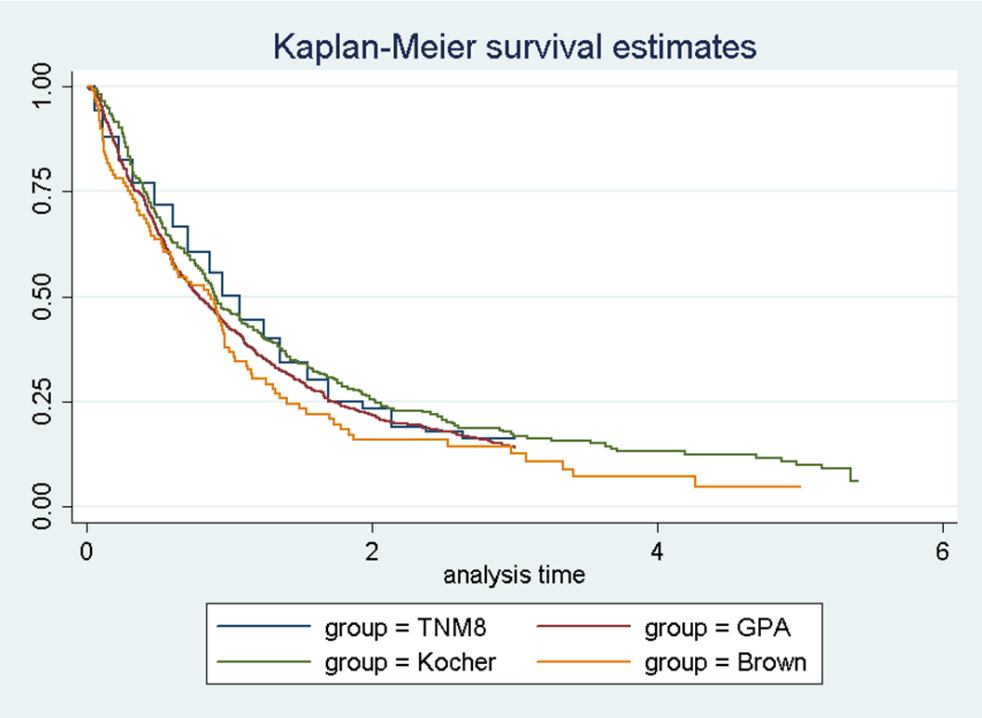


Figure 5: KM Estimates for OS in the TP (1-3) Group

For the purposes of economic modelling, we decided to fit parametric survival models to these KM data because we wanted the curves to be able to work flexibly with a cycle length and time horizon defined by ourselves within the economic model. The best fitting models were selected using Akaike's Information Criterion (AIC). We also restricted our selection to models with a log relative-hazard form rather than an accelerated failure time form. This was because we wanted to use a variety of published hazard ratios to simulate other patient groups within the model. Table 6 shows the AIC statistic was smallest for the Gompertz model in all three datasets

Table 6: AIC Statistics for Log Relative-Hazard Models for OS Curves

	Exp	Weibull	Gompertz
Kocher 2011 OS (No WBRT Arm)	341.47	342.26	339.1295

	Exp	Weibull	Gompertz
Brown 2016 OS (No WBRT Arm)	527.1987	528.8353	524.5493
Sperduto 2016 OS	2356.052	2343.253	2272.63

As per the committee's instructions, we then meta-analysed the shape and scale parameters of the Gompertz curves to obtain the final parameters the curve that represented the OS of patients who were TP (1-3) within the model. In theory it might have been preferable to fit a bivariate model and meta-analysed both the shape and scale parameters together, accounting for correlations, but we felt that independent meta-analyses were reasonable given the small number of studies and the lack of observed correlations between the shape and scale parameters within studies.

Random effects models were chosen due to heterogeneity between the study participants, settings and treatments. The results are in Table 7.

Table 7: Shape and Scale Parameters from the Gompertz Overall Survival Models

Study	Constant mean	Constand SE	Gamma mean	Gamma SE
Brown 2016	0.0435886	0.14708	-0.2411071	0.123695
Kocher 2011	-0.2528468	0.120935	-0.1782433	0.086902
GPA	-0.0682469	0.061599	-0.2512568	0.054425
Meta-analysis	-0.094	0.068	-0.232	0.044

Overall survival curves then needed to be estimated for other groups within the model. It was agreed the best source of evidence for the survival difference between the TP (1-3) group and the TP (4+) group was the hazard ratio of people with 1-4 versus 5+ BM published in the Sperduto study. This hazard ratio came from a multivariable regression so was controlling for a number of other relevant factors and although the difference in the populations is slightly indirect, the committee agreed that it was a reasonable approximation. The Sperduto study publishes separate hazard ratios for people with and without adenocarcinoma histology. We obtained data on the number of patients in our model cohort who were expected to have adeno and non-adeno histology and weighted the hazard ratio accordingly. Separate scenario analyses for these two population groups were also conducted. The hazard ratio obtained from an earlier GPA paper by Sperduto that related to the difference in OS between two broad GPA groups that were representative of the difference between 1-4 and 4+ metastases was also obtained via digitising the relevant survival curves and used in sensitivity analysis.

Table 8: Hazard Ratios and acceleration factors for OS and PFS used within the Model

	mean	Upper CI	Lower CI	Source
Hazard ratio between (1-3) and 4+ group (adeno)	1.28	1.09	1.52	Sperduto 2017
Hazard ratio between (1-3) and 4+ group (non-adeno)	1.03	0.78	1.37	Sperduto 2017
Hazard ratio between (1-3) and 4+ group (GPA)	1.49	1.33	1.66	GPA Sperduto 2012
% of cases adenocarcinoma	0.52	0.40	0.65	NLCA Annual report 2017[m]
Weighted Average Hazard Ratio	1.16	0.90	1.46	Calculated
Hazard ratio for FN4+ vs TP4+	1.21	0.97	1.5	Mulvenna 2016
Acceleration factor for FN(1-3) PFS curve	10%	2.6%	22%	Developer Calculation
Acceleration factor for FN(4+) PFS curve	16.5%	8%	27%	Developer Calculation

Due to the lack of directly relevant data, estimating the OS curves for False Negative patients required some further assumptions, which were discussed in the Model Structure section. OS for patients who were FN (4+) was modelled as being equal to patients who were TP (4+). This was because the committee were unaware of any evidence that earlier detection would significantly affect OS in this group. People in this group were assumed not to be indicated for any radical therapy to their brain and the effect of WBRT on OS is uncertain. The hazard ratio for patients who were FN (1-3) versus TP (1-3) was assumed to begin at 1 at the beginning of the model and progress uniformly, cycle by cycle, to 1.16 (see Table 8) over the 2* median time to intracranial progression observed in the Brown 2016 trial, which was 21 weeks. By week 42, the HR for this group was therefore equal to the group with 4+ BM as it was assumed that the vast majority of the patients would have intracranially progressed. These assumptions were tested in sensitivity analyses.

[m] National Lung Cancer Audit (2018). NLCA annual report 2017. [online] Available at: https://www.rcplondon.ac.uk/projects/outputs/nlca-annual-report-2017 [Accessed 7 Aug. 2018].

Progression Free Survival

The same Kocher 2011 and Brown 2016 BM trials that provided data on OS also provided data on PFS. For Kocher 2011 we obtained the PFS curves through a personal communication with the trialists[n]. The committee were shown both survival curves and concluded that the Kocher 2011 PFS data (again, the no WBRT) arm should be used to model PFS in the base case for people who were TP (1-3) because it showed both intra and extracranial progression and was conducted in a European setting. We digitised the PFS curves from Kocher and Brown and fitted parametric survival models to them via the method described in the Overall Survival Section.

Table 9: AIC Statistics for Parametric Survival Curves fit to PFS Data

AICs for PFS Curves	Exponential	Weibull	Gompertz	Lognormal	Loglogistic
Kocher 2011	595	596	563	553	540
Brown 2016	205	207	203	198	202

Based on the AIC statistics shown in Table 9, we selected a log logistic form for the Kocher data and a lognormal form for the Brown data. In order for the Kocher PFS curve to interact properly with the OS curves within the model we set up the model so that it calculated, cycle-by-cycle, the people alive and progression-free as a proportion of those alive as dictated by the Kocher OS curve. This gave us a 'PFS multiplier' curve that we could then use with other survival curves. The result of this is that, whichever OS curve is used (Kocher, Brown, meta-analysed curve, adeno only e.g.), the proportion of people alive and progression free will remain constant, even though the raw number will change.

The committee considered whether the PFS curves should be meaningfully altered for FN patients to reflect the lack of management that they receive and concluded that they should be. The method for doing this for the FN (1-3) population has already been described in the Model Structure section and details the process by which we arrived at an acceleration factor of 11% during the time that these patients remain undiagnosed. For the FN (4+) patients who would have been treated with WBRT, had they been identified at initial imaging, we calculated an acceleration factor by fitting a loglogistic regression to both arms of the Kocher 2011 PFS data with the study arm representing 'no WBRT' as an independent variable. The acceleration factor associated with this variable was 30.4% (s.e. 11.4%, p=0.001).

[n] EORTC Data Centre (2018) Personal Communication with NICE Centre for Guidelines

Progression and Death Events

Progression is an important concept to capture in NSCLC models because it often triggers challenge of the cancer with another or repeat of therapy. Such therapies are typically of defined and relatively short duration such as 10 sessions of WBRT or 4 cycles of SACT.

The implementation of progression cost within the model is somewhat complex. As partitioned survival analyses are state membership rather than state transition models, there are no transition probabilities between the progression free and progressed health states so these have to be estimated. In our model, these data are only important for cost accrual.

A one-off cost of death was applied by calculating the difference in the overall survival curve (people in the dead state) from cycle to cycle. It is not possible to use this same logic to calculate the number of progressions from the progression free to the progressed state because some of these progressions are deaths. Similarly, one cannot easily treat deaths from the progressed state any differently to deaths from the progression free state without making some assumptions. Our model assumes they had a homogenous cost although this might not be true in reality. This limitation was assessed as minor because the overall proportion of progressions that were deaths was very similar across strategies.

The committee indicated to us that they expected half of FN patients to present with mild to moderate symptoms to their cancer nurse. Upon presentation these patients would undergo imaging, at which point their BM would be discovered. The other half of FN patients were expected to present as an emergency with severe symptoms, resulting in an A&E visit, a non-elective inpatient stay and the requisite imaging.

It was not straightforward to determine what treatments the different populations in the model would receive when experiencing the various events in the progression decision trees (see paragraph below) and we had no evidence to inform these parameters other than committee assumption. Firstly, we needed to determine which False Negative (1-3) patients would still receive radical brain treatment upon intracranial progression. Since we assumed that 50% of people would progress as a routine presentation with mild symptoms, the committee agreed that it would be reasonable to assume that 50% of patients would receive radical brain treatment if intracranial progression was part of their first event (whether alone or along with extracranial progression). This assumption could be changed to apply to only patients whose first event was intracranial alone or who experienced any intracranial event. 80% of patients who were FN (4+) were assumed to receive SACT upon intracranial progression (the same proportion as if they had been identified early). Underlying intracranial progression event costs that applied to all patients were also applied; 80% received WBRT, 5% SRS and 5% SACT. For the patients who were TP (4+), the WBRT was removed as they had received this intervention on initial diagnosis.

To calculate the weighted average cost of a progression event we obtained the progression event decision trees (see Table 33 for those data) from the Kocher 2011 trial for patients with initial treatment with WBRT (TP4+) and without WBRT (all other patients). Deaths were assigned a cost of £0 because they are already accounted for via the method detailed above. Those who did not progress at all were removed from the decision tree because they are not relevant to the calculation.

60% of patients who progressed extracranially alone first were assumed to receive SACT. 20% of these patients as well as any patients who had intracranial and extracranial progression, whether together or consecutively in any order were assumed to receive a single dose of palliative radiotherapy.

All treatment assumptions were provided by the committee. The constituent and resulting cost data are provided from Table 28 onwards. Death costs are available in Table 22.

State Membership Costs

The longer term partitioned survival analysis model contains three possible membership states for simulated patients; progression free survival, progressed and dead. Patients in each of these states consume resources at differing amounts, and therefore incur differing total costs for each given unit of time they have membership of the states.

In order to arrive at state membership costs for the aforementioned states, we examined the literature to uncover the types of resource that commonly were used in each membership state, and the associated numbers of units consumed each month. We developed this information into a table and presented it to the committee alongside up to date prices for resource units from the English NHS. The committee used this table as a starting point to a discussion to validate these data for use within the economic model. The committee made changes to this table based on their experience of the NHS, excising some resource use, unit usage and costs, whilst adding others. The committee also agreed that the state membership costs were the same, despite the stage of cancer the patient experiences.

The committee agreed that patients stop incurring ongoing costs when they die.

Here we present tables to show the final membership costs of progression free survival (Table 10), progressed (Table 11), agreed by the committee to be valid for use in the economic model:

Table 10. Long term model - Progression free survival membership

Resource type	Percentage of patients who use the resource each month (committee assumption)	Number of units used per patient each month (committee assumption)	Unit cost (SE)	Reference for unit cost
Hospitalisation	2.5%	1	£1,590.00 (£397.50)	NHS National Schedule of Reference Cost 2016/17[o]
Cancer Nurse	70.0%	1	£38.75 (£0.02)	NHS National Schedule of Reference Cost 2016/17 - N21AN
Outpatient (Multi-professional Non-Admitted Face-to-Face Attendance, Follow-up)	75.0%	1	£191.11 (£0.45)	NHS National Schedule of Reference Cost 2016/17 - WF02A
GP Visit	10.0%	1	£38.00 (£9.50)	PSSRU 2017 [p] General Practitioner - per patient contact lasting 9.22 minutes Including direct care staff costs, and qualifications costs, p162
Complete blood count	100.0%	0.75	£3.06 (£0.00)	NHS National Schedule of Reference Cost 2016/17 - DAPS05
Palliative radiotherapy	12.5%	1	£132.40 (£33.10)	NHS National Schedule of Reference Cost 2016/17 – Same as SC23Z
CT scan	60.0%	0.75	£120.07 (£0.16)	NHS National Schedule of Reference Cost 2016/17 - RD22Z
X-Ray of chest	100.0%	0.333	£25.00	FOI Request (23023) Stockport NHS Trust 2014
Biochemistry	100.0%	0.75	£1.13 (£0.00)	NHS National Schedule of Reference Cost 2016/17 - DAPS04

[o] Improvement.nhs.uk. (2018). Reference costs | NHS Improvement. [online] Available at: https://improvement.nhs.uk/resources/reference-costs/ [Accessed 6 Aug. 2018].

[p] Curtis, Lesley A. and Burns, Amanda (2017) Unit Costs of Health and Social Care 2017. Report number: https://doi.org/10.22024/UniKent/01.02/65559. Personal Social Services Research Unit, University of Kent, 260 pp. ISBN 978-1-911353-04-1. (doi:https://doi.org/10.22024/UniKent/01.02/65559) (Full text available)

In order to arrive at the costs for each patient for each month whilst they have membership of the progression free survival state, we multiplied the percentage of patients who are assumed to use the resource type each month, by the number of units used by those patients, by the unit cost to obtain the total weighted cost. For progression free survival patients this was £296.06. We then multiplied this value by the number of months in a year (12) and divided by the number of cycles the model uses each year (52) to obtain a progression free survival cycle cost of £68.32.

Table 11. Long term model - Progression membership

Resource type	Percentage of patients who use the resource each month (committee assumption)	Number of units used per patient each month (committee assumption)	Unit cost (SE)	Reference for unit cost
Hospitalisation	20.00%	1	£1,590.00 (£397.50)	NHS National Schedule of Reference Cost 2016/17
Cancer Nurse	10.00%	1	£38.75 (£0.02)	NHS National Schedule of Reference Cost 2016/17 - N21AN
Palliative Care Nurse	20.00%	1	£102.41 (£0.50)	NHS National Schedule of Reference Cost 2016/17 - N21AF
Palliative Care Physician	80.00%	2	£158.81 (£39.70)	NHS National Schedule of Reference Cost 2016/17 - SD04A
Outpatient (Multi-professional Non-Admitted Face-to-Face Attendance, Follow-up)	100.00%	1	£191.11 (£0.45)	NHS National Schedule of Reference Cost 2016/17 - WF02A
GP Visit	28.00%	1	£38.00 (£9.50)	PSSRU 2017 General Practitioner - per patient contact lasting 9.22 minutes Including direct care staff costs, and qualifications costs, p162
Stereoids (Dexamethasone 0.5mg tablets)	50.00%	16	£0.58 (£0.14)	Drug Tariff (May 2018)[q]

[q] Drugtariff.nhsbsa.nhs.uk. (2018). NHS Electronic Drug Tariff. [online] Available at: http://www.drugtariff.nhsbsa.nhs.uk/#/00446515-DC_2/DC00446511/Home [Accessed 14 May 2018].

Resource type	Percentage of patients who use the resource each month (committee assumption)	Number of units used per patient each month (committee assumption)	Unit cost (SE)	Reference for unit cost
NSAIDS (ibuprofen 200mg tablets)	30.00%	60	£0.03 (£0.01)	Drug Tariff (May 2018)
Morphine (20mg tablets)	75.00%	21	£0.19 (£0.05)	Drug Tariff (May 2018)
Complete blood count	100.00%	1	£3.06 (£0.00)	NHS National Schedule of Reference Cost 2016/17 - DAPS05
Palliative radiotherapy	20.00%	1	£132.40 (£33.10)	NHS National Schedule of Reference Cost 2016/17 – Same as SC23Z
Biochemistry	100.00%	1	£1.13 (£0.00)	NHS National Schedule of Reference Cost 2016/17 - DAPS04
CT scan	70.00%	0.333	£120.07 (£0.16)	NHS National Schedule of Reference Cost 2016/17 - RD22Z
Home oxygen	5.00%	7	£107.84	http://www.emrespiratory.co.uk/downloads/documents/HOSAR-Good-Practice-Guide.pdf
X-Ray	30.00%	0.75	£25.00	FOI Request (23023) Stockport NHS Trust 2014
Anti-epileptics (Levetiracetam 250mg x 60)	77.1%	1	£19.31	Drug Tariff (May 2018)

In order to arrive at the costs for each patient for each month whilst they have membership of the progressed state, used the same approach as the progression free survival state. The resulting figures are a weighted average progressed state membership cost of £923.24 each month, and a cycle cost of £213.06.

Initial treatments

The committee were consulted on the types of treatments that patients would be eligible to receive, and what percentage of patients eligible would receive them, given the number of brain metastases detected by the initial diagnostic strategy. The committee were also consulted with regards to the costs of such treatments. Here we present how we calculated the costs for each of the treatments used in the model, all of which were validated by the committee.

Surgical treatments for primary tumours

Table 12 shows the costs of surgical procedures for primary tumours in patients with lung cancer. There are no reference costs that apply to the specific treatments listed so the committee chose the most appropriate from the full range of available thoracic procedure reference costs. The cost of 'Complex resections and other resections' was calculated by averaging the cost of lobectomy and pneumonectomy.

Table 12. Surgical procedure for primary tumour

Type of treatment	Cost of treatment (SE)	Reference for treatment cost
Lobectomy	£6,522.66 (£31.79)	NHS National Schedule of Reference Cost 2016/17 - DZ02K
Wedge resection	£3,595.15 (£40.45)	NHS National Schedule of Reference Cost 2016/17 - DZ64B
Pneumonectomy	£7,562.42 (£42.72)	NHS National Schedule of Reference Cost 2016/17 - DZ02J
Complex resections and other resections	£7,042.54	Average cost of Lobectomy and Pneumonectomy

Radiotherapy treatments for primary tumours

Stereotactic Ablative Radiotherapy (SABR)

Stereotactic Ablative Radiotherapy (SABR), is an emerging technology. It is a specialised radiotherapy treatment planning technique resulting in a high dose to the target with steep dose gradients resulting in rapid dose fall off outside the target area. This results in high biologically effective dose (BED) while minimising the dose received by the normal tissues, and could potentially minimise the radiotherapy treatment toxicity and side effects. SABR is currently provisioned by the NHS through the Commissioning through Evaluation (CtE) programme, whilst it awaits a full formal review for general use in the NHS. The CtE tariff (Table 13) reimburses three different treatment regimens, 3 fractions, 5 fractions and 8 fractions. These tariffs have been identified by Leeds Teaching Hospital as bundled tariffs, meaning that they include payments for all related planning and treatment.

Table 13. SABR tariff

Regimen	NHSE Tariff 2017/2018 (SE)	Proportion of patients who receive each treatment regimen	Reference
SABR CtE - 3 fractions	£3,574.99 (£893.75)	0.165	Leeds Teaching Hospital, NHS Trust
SABR CtE - 5 fractions	£5,058.76 (£1,264.69)	0.671	Leeds Teaching Hospital, NHS Trust
SABR CtE - 8 fractions	£7,283.42 (£1,820.86)	0.164	Leeds Teaching Hospital, NHS Trust

In order to obtain the cost of SABR for an average patients, the tariff costs must be weighted by the proportion of patients receiving each regimen. This information was provided by Leeds Teaching Hospital, NHS Trust. When the costs of each regimen are weighted against the proportion of patients who receive each treatment regimen, the average cost of SABR for a patient is calculated to be £5,178.78. The costs of SABR are expected to decline with routine adoption.

From the NLCA data, we find that overall, for stage I and II NSCLC, 53.9% of patients receive SABR. Using this, an assumption made by the committee that a patient would be twice as likely to receive SABR in stage I as stage II disease, and the data found in

Table 14, we calculate that 63.38% of patients in stage I and 31.69% of patients with stage II NSCLC receive SABR.

Table 14. Patients who presented with NSCLC from the NLCA Report 2017

Regimen	Percentage of all patients who presented	Stage total	Reference
Stage Ia	12.0%	19%	NLCA Report 2017
Stage Ib	7.0%		NLCA Report 2017
Stage IIa	4.0%	8%	NLCA Report 2017
Stage IIb	4.0%		NLCA Report 2017

Continuous hyperfractionated accelerated radiotherapy (CHART)

Continuous hyperfractionated accelerated radiotherapy (CHART) is a method of delivering standard external beam radiotherapy in a more intense regimen than conventional radiotherapy. The CHART regimen used in the model assumes 55Gy delivered over 36 sessions over 12 days, including weekends.

Table 15. CHART for primary tumour

Resource type	Number of resource units used	Resource unit cost (SE)	Reference
Define volume for simple radiation therapy with imaging and dosimetry	1	£362.59 (£1.31)	Unit cost from NHS National Schedule of Reference Cost 2016/17 - SC45Z Resource use from CG121
Deliver a fraction of complex treatment on a megavoltage machine	1	£132.40 (£0.04)	Unit cost from NHS National Schedule of Reference Cost 2016/17 - SC23Z Resource use from CG121
Deliver a fraction of treatment on a megavoltage machine	35	£107.46 (£0.10)	Unit cost from NHS National Schedule of Reference Cost 2016/17 - SC22Z Resource use from CG121
Number of days of hospital inpatient stay	12	First 5 days - £1,590 (£397.50) Excess bed days - £313	NHS National Schedule of Reference Cost 2016/17

To calculate the total cost of CHART, the number of resource units used is multiplied by the resource unit cost. The cost of hospital inpatient stay is calculated as the initial cost of first 5 days stay (£1,590) added to the remainder of hospital inpatient stay days (12-5) multiplied by the cost of excess bed days (£313). When these costs are added together, this results in the total cost of CHART for each patient as £8,037.25.

Hyper fractionated accelerated radiotherapy

Hyper fractionated accelerated radiotherapy in our model was defined as the delivery of 55Gy over 20 sessions over the course of four weeks. Table 15 shows the how the cost of hyper fractionated accelerated radiotherapy was calculated. This is the most common form of radical radiotherapy practiced in the UK NHS today.

Table 16. Hyper fractionated accelerated radiotherapy

Resource type	Number of resource units used	Resource unit cost (SE)	Reference
Define volume for simple radiation therapy with imaging and dosimetry	1	£362.59 (£1.31)	Unit cost from NHS National Schedule of Reference Cost 2016/17 - SC45Z Resource use from CG121
Deliver a fraction of complex treatment on a megavoltage machine	1	£132.40 (£0.04)	Unit cost from NHS National Schedule of Reference Cost 2016/17 - SC23Z Resource use from CG121
Deliver a fraction of treatment on a megavoltage machine	19	£107.46 (£0.10)	Unit cost from NHS National Schedule of Reference Cost 2016/17 - SC22Z Resource use from CG121

To calculate the cost of hyper fractionated accelerated radiotherapy, we multiply the number of resource units by the cost of each unit, and add them together. This results in the cost of hyper fractionated accelerated radiotherapy for each patient at £2,536.81.

Standard fractionated radiotherapy

Standard fractionated radiotherapy in our model was defined as the delivery of 60-66 Gy over 30-33 sessions over the course of 6 – 6.5 weeks. Table 17 shows the how the cost of standard fractionated accelerated radiotherapy was calculated.

Table 17. Standard fractionated radiotherapy

Resource type	Number of resource units used	Resource unit cost (SE)	Reference
Define volume for simple radiation therapy with imaging and dosimetry	1	£362.59 (£1.31)	Unit cost from NHS National Schedule of Reference Cost 2016/17 - SC45Z Resource use from CG121
Deliver a fraction of complex treatment on a megavoltage machine	1	£132.40 (£0.04)	Unit cost from NHS National Schedule of Reference Cost 2016/17 - SC23Z Resource use from CG121
Deliver a fraction of treatment on a megavoltage machine	29	£107.46 (£0.10)	Unit cost from NHS National Schedule of Reference Cost 2016/17 - SC22Z Resource use from CG121

To calculate the cost of hyper fractionated accelerated radiotherapy, we multiply the number of resource units by the cost of each unit, and add them together. This results in the cost of standard fractionated radiotherapy for each patient at £3,611.46.

Fractionated radiotherapy for local control – 36 Gy over 12 sessions

The costing for fractionated radiotherapy for local control – 36 Gy over 12 sessions, is shown in Table 18. The total cost for fractionated radiotherapy for local control – 36 Gy over 12 sessions was found to be £1,652.16.

Table 18. Fractionated radiotherapy for local control 36 Gy over 12 sessions

Resource type	Number of resource units used	Resource unit cost (SE)	Reference
Define volume for simple radiation therapy with imaging and dosimetry	1	£362.59 (£1.31)	Unit cost from NHS National Schedule of Reference Cost 2016/17 - SC45Z Resource use guideline committee
Deliver a fraction of treatment on a megavoltage machine	12	£107.46 (£0.10)	Unit cost from NHS National Schedule of Reference Cost 2016/17 - SC22Z Resource use guideline committee

Fractionated radiotherapy for local control – 20 Gy over 5 sessions

The costing for fractionated radiotherapy for local control – 20 Gy over 5 sessions, is shown in Table 19. The total cost for fractionated radiotherapy for local control – 20 Gy over 5 sessions was found to be £899.91.

Table 19. Fractionated radiotherapy for local control 20 Gy over 5 sessions

Resource type	Number of resource units used	Resource unit cost (SE)	Reference
Define volume for simple radiation therapy with imaging and dosimetry	1	£362.59 (£1.31)	Unit cost from NHS National Schedule of Reference Cost 2016/17 - SC45Z Resource use guideline committee
Deliver a fraction of treatment on a megavoltage machine	5	£107.46 (£0.10)	Unit cost from NHS National Schedule of Reference Cost 2016/17 - SC22Z Resource use guideline committee

Radiotherapy for local control is given to some stage IIIA patients who are positive for brain metastases within the model.

Treatments for brain tumours

Stereotactic radiosurgery

The cost of stereotactic radiosurgery, £3,555.65, was taken from the model which was created for NICE Guideline NG99 (Brain tumours (primary) and brain metastases in adults). As the NICE Brain Tumour model did not specify a standard deviation for the cost of stereotactic radiosurgery, we assumed this to be a quarter of the mean price (£888.91).

Surgical brain resection

The cost of surgical brain resection, £7,031.94, was taken from NICE Guideline NG99 (Brain tumours (primary) and brain metastases in adults). As the guideline did not specify a standard deviation for the cost of surgical brain resection, we assumed this to be a quarter of the mean price (£1,757.98).

Whole brain radiotherapy (WBRT)

Whole brain radiotherapy (WBRT) included in our model consisted of preparation 10 fractions.

Table 20. Whole brain radiotherapy

Resource type	Number of resource units used	Resource unit cost (SE)	Reference
Preparation of simple radiotherapy with imaging and dosimetry, with technical support	1	£449.70 (£5.39)	Resource use from - Addenbrookes Hospital NHS National Schedule of Reference Cost 2016/17 - SC46Z
Deliver a fraction of treatment on a megavoltage machine cost	10	£107.46 (£0.10)	Resource use from - Addenbrookes Hospital NHS National Schedule of Reference Cost 2016/17 - SC22Z

To calculate the cost of WBRT, we multiply the number of resource units by the cost of each unit, and add them together. This results in the cost of WBRT at £1,524.34.

Systemic Anti-Cancer Therapy (SACT)

There are a very large number of systemic therapy options available in NSCLC (see RQ 3.3 of this update for a full algorithm) so costing them all and factoring in their differential benefits in this patient population would have been impractical and subject to high uncertainty. These treatment options have typically been the subject of NICE Technology Appraisals and therefore represent cost-effective additions to the care pathway, but additions that the committee was aware were unlikely to add much in terms of net monetary benefit. This is because Technology Appraisal approved drugs in advanced cancer rarely have base case ICERs significantly lower than the upper limit of the ICER range normally considered cost-effective by NICE. The committee also noted that much of the evidence in this model came from survival data collected before many of these drugs were widely available. They therefore thought that the net monetary benefit associated with systemic therapy could reasonably be approximated using the costs of a representative platinum doublet chemotherapy. Systemic anti-cancer therapy (SACT) treatment in our model therefore consisted of Vinorelbine (oral), Carboplatin (IV), and Dexamethasone (oral). In the base case, patients received 4 cycles for each course of SACT. Each course of SACT required a quarter of an hour of an Agenda for Change band 4 member of staff to book an outpatient appointment.

The dose of oral Vinorelbine required for patients is 60mg/mg^2, which equates to 120mg on days 1 and days 8 of each cycle. We assumed that the Carboplatin dose required equated to a target AUC 5mg/ml/min, based on a surface area of 1.73m2 and an eGFR of 90. This translated to a requirement of 575mg of Carboplatin required for infusion each cycle. The dosage regimen of dexamethasone was calculated based on the advice of the guideline committee as 8mg twice a day over the first week, tapering down over the remaining 3 weeks.

Table 21. Systemic Anti-Cancer Therapy

Resource type	Number of resource units used per cycle	Resource unit cost (SE)	Reference
Administration Outpatient Appointment Booking - AfC Band 4 hourly rate	0.25	£28 (£0.13)	PSSRU 2017
Outpatient appointment	1	£173.99	NHS National Schedule of Reference Cost 2016/17 - SB12Z
Vinorelbine 20mg (oral capsules)	4	£43.98	BNF Online [Accessed 19th July 2018]

Resource type	Number of resource units used per cycle	Resource unit cost (SE)	Reference
Vinorelbine 80mg (oral capsules)	2	£175.50	BNF Online [Accessed 19th July 2018]
Carboplatin 150mg/15ml solution for infusion vials	3.833	£6.35	eMIT National 2016/2017 NCP Code DHE001[r]
Dexamethasone 0.5mg – Box of 28 tablets	2.9	£14.25	Drug Tariff May 2018
Dexamethasone 2mg – Box of 50 tablets	2.035	£16.22	Drug Tariff May 2018

The sum of resource use in Table 21 summates to the cost of each SACT cycle as £750.84. Therefore, the cost of all 4 cycles is £3,003.36.

Death event

To calculate the cost of a death event in the mode, we used resource costs from Georghiou and Bardsley (2014), given over to the patient in the final three months of their lives. From this, study, we sum the average hospital costs, local authority funded care, district nursing care, GP contacts costs and inflate them to 2018 levels using a four yearly inflation factor of ~6% (PSSRU HCHS). As patients accrue the death event costs during the final three months of their lives, we account for this by removing the state based costs incurred by these patients for being in the model for 3 months with health states weighted by the proportion of people who die directly from the progression free and progressed states.

Table 22. Death event costs

Resource type	Resource unit cost	Reference
Hospital Costs	£5,890.00	Developer assumption
Local Authority Funded Care	£444.00	Developer assumption
District Nursing Care	£588.00	Developer assumption
GP Contacts	£365.00	Developer assumption

This results in the death event total cost (less the weighted state membership costs) to be £5,152.88 (SE £1,288.22).

[r] GOV.UK. (2018). Drugs and pharmaceutical electronic market information tool (eMIT). [online] Available at: https://www.gov.uk/government/publications/drugs-and-pharmaceutical-electronic-market-information-emit [Accessed 7 Aug. 2018].

Patients groups considered by the model

The model considers treatment strategies for stage I, stage II and stage IIIA patients. The stage IIIA patient group consist of five broad treatment strategies; those treated with Chemotherapy and Surgery (CS), Chemotherapy and Radiotherapy (CR), Chemotherapy, Radiotherapy and Surgery (CRS), Radiotherapy only (R) and Surgery only (S). The committee agreed that if they were to deliberate a separate recommendation for each of these five identified treatment strategies for stage III, the resulting guidance would be impractical. Therefore, we have combined and weighed each of the treatment strategies for stage IIIA patients into a single treatment strategy within the model.

Table 23. Treatment strategy split for stage IIIA NSCLC patients

Treatment strategy for stage IIIA patients	% of all patients	Reference
Chemotherapy and Surgery (CS)	26.2%	NLCA Annual report 2017
Chemotherapy and Radiotherapy (CR)	29.1%	NLCA Annual report 2017
Chemotherapy, Radiotherapy and Surgery (CRS)	5.2%	NLCA Annual report 2017
Radiotherapy only (R)	19.2%	NLCA Annual report 2017
Surgery only (S)	20.4%	NLCA Annual report 2017

As it was not directly reported in the NLCA Annual report, the committee advised that only roughly one out of six patients who received chemotherapy and surgery would also receive radiotherapy. Using this information in combination with the data from the National Lung Cancer Audit (NLCA) Report 2017, we calculated the percentage of patients who receive each treatment strategy (shown in Table 23).

Initial Treatments for False Negatives

Whilst the ‘no imaging’ strategies and both imaging strategies result in false negative patients, with between one and three brain metastases, only the ‘no imaging’ strategy result in false negative patients with more than three brain metastases. Since there is no way to distinguish false negatives from true negatives, false negative patients continue to receive the planned initial radical treatment.

The committee agreed that the split between patients who received each treatment for their primary tumour was the same for both stage I and stage II lung cancer patients. As discussed above, patients receiving each type of treatment for stage IIIA lung cancer were weighted into a single model arm.

Here, in Table 24, we present the initial treatment strategies for false negative patients, as taken from the NLCA Annual report 2017 and confirmed by the committee for each aforementioned group.

Table 24. Treatment strategies for patients with undetected brain metastases (False Negative) by cancer stage

	Stage of Lung Cancer diagnosed at the time of imaging strategy							
Treatment strategies + percentage of patients eligible for each treatment	**Stage I**	**Stage II**	**Stage IIIA weighted**	**Stage IIIA – Chemotherapy + Surgery**	**Stage IIIA – Chemotherapy + Radiotherapy**	**Stage IIIA-N2 – Chemotherapy + Radiotherapy + Surgery**	**Stage IIIIA – Radiotherapy**	**Stage IIIA – Surgery**
% patients operable	75.4%	75.4%	47.6%	87.4%	0%	83.5%	0%	100%
Lobectomy	75.7%	75.7%	93.8%	93.8%	0%	0%	0%	93.8%
Wedge resection	17.6%	17.6%	0%	0%	0%	0%	0%	0%
Pneumonectomy	5.0%	5.0%	6.3%	6.3%	0%	100%	0%	6.3%
Complex resections and other resections	1.7%	1.7%	0%	0%	0%	0%	0%	0%
% patients Radiotherapy	24.6%	24.6%	53.5%	0%	100%	100%	100%	0%
SABR	63.4%	31.7%	0%	0%	0%	0%	50%	0%
CHART (55 Gy/5#/1.5 weeks)	3.0%	3.0%	4.4%	0%	4.4%	0%	0%	0%
Standard Fractionated Radiotherapy 60–66 Gy/30–33#/6–6.5 weeks	3.0%	3.0%	4.4%	0%	4.4%	0%	0%	0%
Hypofractionated Radiotherapy 55 Gy/20#/4 weeks	30.6%	62.3%	91.2%	0%	91.2%	100%	50%	0%
% patients Systemic Anti-Cancer Therapy (SACT)	0.0%	0.0%	60.5%	100%	100%	100%	0%	0%

Initial Treatments for True Positives (1-3)

Of the three strategies considered by the model, No Imaging, CT followed by MRI, and MRI alone, only the latter two diagnostic strategies are able to confirm the presence of any number of a brain metastases. Table 25 shows the committee consensus for what treatments would be given to those with 1-3 detected brain metastases and treatments would be given to those eligible to receive radical treatment therapy.

Table 25. Treatment strategies for patients with 1-3 brain metastases (true positive) by cancer stage

	Stage of Lung Cancer diagnosed at the time of imaging strategy							
Treatment strategies + percentage of patients eligible for each treatment	**Stage I**	**Stage II**	**Stage IIIA weighted**	**Stage IIIA – Chemotherapy + Surgery**	**Stage IIIA – Chemotherapy + Radiotherapy**	**Stage IIIA-N2 – Chemotherapy + Radiotherapy + Surgery**	**Stage IIIA – Radiotherapy**	**Stage IIIA – Surgery**
Brain metastases treatment Stereotactic radiosurgery	75%	75%	10%	10%	10%	10%	10%	10%
Brain metastases treatment Surgical brain resection	10%	10%	0%	0%	0%	0%	0%	0%
Brain metastases treatment WBRT	10%	10%	0%	0%	0%	0%	0%	0%
Brain metastases treatment No treatment	5%	5%	0%	0%	0%	0%	0%	0%
Brain metastases treatment SACT	80%	80%	100%	100%	100%	100%	100%	100%
Local control Radiotherapy	10%	10%	40%	40%	40%	40%	40%	40%
% patients treatments for radical treatment	75%	75%	0%	0%	0%	0%	0%	0%
% radical treatments that are surgery	20%	20%	0.%	0%	0%	0%	0%	0%
% radical treatments that are radiotherapy	80%	80%	74.4%	100%	100%	0%	100%	0%

Initial Treatments for True Positives (4+)

Table 26 shows the committee consensus for what treatments would be given to those with more than 3 detected brain metastases. The committee assumed that 15% of patients with more four or more detected brain metastases receive radiotherapy for local control, 92.5% would receive WBRT and 80% of stage I and II patients would receive SACT, with 100% of stage IIIA patients receiving SACT. In our model, patients with more than 3 brain metastases do not receive any radical therapy.

Table 26. Treatment strategies for patients with more than 3 brain metastases (true positive) by cancer stage

	Stage of Lung Cancer diagnosed at the time of imaging strategy							
Treatment strategies + percentage of patients eligible for each treatment	**Stage I**	**Stage II**	**Stage IIIA weighted**	**Stage IIIA – Chemotherapy + Surgery**	**Stage IIIA – Chemotherapy + Radiotherapy**	**Stage IIIA-N2 – Chemotherapy + Radiotherapy + Surgery**	**Stage IIIA – Radiotherapy**	**Stage IIIA – Surgery**
Brain metastases treatment SACT	80%	80%	100%	100%	100%	100%	100%	100%
Brain metastases treatment WBRT	92.5%	92.5%	92.5%	92.5%	92.5%	92.5%	92.5%	92.5%
Local control Radiotherapy	15%	15%	15%	15%	15%	15%	15%	15%
% patients who receive radical treatment	0%	0%	0%	0%	0%	0%	0%	0%

Radiotherapy for local control

Patients with any number of brain metastases may receive radiotherapy for local control, as indicated in Table 25 or Table 26. Where this is the case, 25% of patients who receive radiotherapy for local control receive 36 Gy over 12 sessions, whilst the remaining 75% of patients receive 20 Gy over 5 sessions.

Initial Imaging Strategies

As described earlier, received either an MRI scan alone, or a CT scan, followed by a confirmatory MRI scan, or no imaging.

Table 27 the costs of imaging modalities used in the model.

Table 27. Imaging strategy costs

Imaging strategy	Cost of strategy (SE)	Reference
CT scan	£120.07 (£0.16)	NHS National Schedule of Reference Cost 2016/17 – RD22Z
MRI scan	£180.48 (£0.26)	NHS National Schedule of Reference Cost 2016/17 – RD03Z

Progression and presentation

As discussed earlier, half of patients who were FN are expected to present as an emergency with severe symptoms, whilst the other half are expected to present in a routine appointment with their cancer nurse after experiencing mild symptoms. In the model, both of these types of presentation are associated with significantly different resource use and associated cost.

Here in Table 28, we present the cost of emergency presentation and in Table 29 for non-emergency routine presentation for FN patients.

Table 28. FN Emergency presentation resource use and cost

Resource type	Number of resource units used per cycle (committee assumptions)	Resource unit cost (SE)	Reference
A&E	1	£148.00 (£37.00)	PSSRU 2017
Inpatient hospital stay (5 days)	1	£1,590.00 (£397.50)	NHS National Schedule of Reference Cost 2016/17 - SB12Z
CT scan	1	£120.07 (£0.16)	Resource Use – Guideline Committee Cost - NHS National Schedule of Reference Cost 2016/17 – RD22Z
MRI scan	1	£180.48 (£0.26)	Resource Use – Guideline Committee Cost - NHS National Schedule of Reference Cost 2016/17 – RD03Z

Table 29. FN routine presentation resource use and cost

Resource type	Number of resource units used per cycle (committee assumptions)	Resource unit cost (SE)	Reference
Specialist nurse in outpatient clinic	1	£191.11 (£0.45)	PSSRU 2017
CT scan	1	£120.07 (£0.16)	Resource Use – Guideline Committee Cost - NHS National Schedule of Reference Cost 2016/17 – RD22Z
MRI scan	1	£180.48 (£0.26)	Resource Use – Guideline Committee Cost - NHS National Schedule of Reference Cost 2016/17 – RD03Z

Summing the costs gives a total for emergency presentation of £2,038.55 and routine presentation as £491.65. Assuming 50% of intracranial progressions for FN patients are of each type, the average cost in the model is £1,265.10.

Intracranial and extracranial progression event

As noted in the sections on progression above, there are several different types of progression events, including intracranial, extracranial, and both intracranial and extracranial. Each one of these pathways is associated with different levels of resource use and therefore overall cost. Here we present the average cost associated with each type of progression event within the progression decision trees (see below).

Table 30. Intracranial Progression Event Treatment cost

Resource type	Proportion of patients who use the resource	Resource unit cost (SE)	Reference
Whole Brain Radiotherapy (WBRT)	0.8	£1,524.34	Resource Use – Guideline Committee Unit cost - Calculated for this model
Stereotactic radiosurgery (SRS)	0.05	£3,555.65 (£888.91)	Resource Use – Guideline Committee Unit cost - NICE Guideline NG99
SACT (4 cycles)	0.05	£3,003.36	Resource Use – Guideline Committee Unit cost - Calculated for this model

Therefore, we calculate the cost of an intracranial progression event to be £1,547.42 (SE of £386.86).

The additional cost of an Intracranial Progression Event Cost for TP4+ patients is the same as shown in Table 30, except that instead of 80% of patients receiving WBRT, no patients receive WBRT. This results in the cost of an Intracranial Progression Event Cost for TP4+ patients as £327.95.

The additional cost of an Intracranial Progression Event for FN patients with 1-3 brain metastases was calculated to be £4,087.65, which assumes that 50% of patients presenting late will be treated with radical treatment, whilst the cost of an Intracranial Progression Event Cost for FN patients with more than 3 brain metastases was calculated to be £2,402.69, which is simply the cost of SACT multiplied by the assumed probability that those patients would receive it (80%).

Table 31. Extracranial Progression Event Treatment cost

Resource type	Proportion of patients who use the resource	Resource unit cost (SE)	Reference
SACT (4 cycles)	0.6	£3,003.36	Resource Use – Guideline Committee Unit cost - Calculated for this model
Palliative radiotherapy single fraction 1-5	0.2	£132.40 (£33.10)	Resource Use – Guideline Committee Unit cost - NHS National Schedule of Reference Cost 2016/17 – SC22Z

The cost of an extracranial progression event is the sum of these values; £1,828.50 (SE of £457.12).

Table 32. Intracranial and Extracranial Progression Event Treatment cost

Resource type	Proportion of patients who use the resource	Resource unit cost (SE)	Reference
Palliative radiotherapy single fraction 1-5	0.2	£132.40 (£33.10)	Resource Use – Guideline Committee Unit cost - NHS National Schedule of Reference Cost 2016/17 – SC22Z

The cost of an intra and extracranial progression event (whether occurring together or separately) is the sum of these values; £26.48 (SE of £6.62).

Intracranial and extracranial progression event decision tree

The trialists for Kocher 2011 provided additional data of probabilities of progression events after intracranial progression (Table 33).

Table 33. Progression and death event probabilities for patients who are given or not given WBRT

Parameter	No WBRT	WBRT	Reference
Probability of Death before progression	0.091954023	0.149700599	Kocher 2011 Supplementary Data
Probability of Intracranial + Extracranial progression	0.109195402	0.05988024	
Probability of Intracranial progression	0.465517241	0.293413174	
Probability of Extracranial progression	0.333333333	0.497005988	
Probability of Death after Intracranial progression	0.432098765	0.489795918	
Probability of Extracranial progression after Intracranial progression	0.395061728	0.346938776	
Probability of Alive after Intracranial progression	0.172839506	0.163265306	
Probability of Death after Extracranial progression	0.362068966	0.65060241	
Probability of Intracranial progression after Extracranial progression	0.586206897	0.277108434	
Probability of Alive after Extracranial progression	0.051724138	0.072289157	
Probability of Death after Intracranial + Extracranial progression	0.947368421	0.9	
Probability of Death after Extracranial after Intracranial progression	0.875	0.913043478	
Probability of Death after Intracranial after Extracranial progression	0.735294118	0.941176471	

These probabilities were used to calculate the number of patients who would experience each type of progression event and the weighted cost (Table 33).

Table 34. Weighted cost of a progression event for each type of patient in the model

Parameter	Cost
Weighted average cost of a progression event (TP 1-3):	£1,342.79
Weighted average cost of a progression event TP (4+):	£1,012.93
Weighted average cost of a progression event (FN 1-3):	£4,840.78
Weighted average cost of a progression event (Undetected 4+):	£3,872.41

Table 34 shows the final weighted cost of a progression event that is arrived at under the base case assumptions in the model.

Utilities

The three health states in the long-term model are associated with utility scores, which are shown in Table 35. Patients who spend time in one or more of these states in the long-term model accumulate QALYs. A final modifying factor for the total number of QALYs a patient may accumulate is the QALY loss associated with surgery. In the base case the data from Lester-Coll 2016 were used for the progression-free and post-progression survival states with the Nafees 2008 data being used in sensitivity analysis.

Table 35. Utilities in the long-term model

Utilities	Utility score (SD)	Reference
HRQoL of Progression-Free Lester-Coll 2016 (SRS)	0.8 (0.12)	Lester-Coll 2016[s]
HRQoL of Progression-Free Nafees 2008 (Stable disease)	0.6532	Nafees 2008[t]
HRQoL of Progressed Lester-Coll 2016 (WBRT)	0.54 (0.15)	Lester-Coll 2016
HRQoL of Progressed Nafees 2008 (Progressive disease adjust)	-0.1798	Nafees 2008

[s] Lester-Coll, Nataniel H., Charles E. Rutter, Trevor J. Bledsoe, Sarah B. Goldberg, Roy H. Decker, and B. Yu James. "Cost-effectiveness of surgery, stereotactic body radiation therapy, and systemic therapy for pulmonary oligometastases." International Journal of Radiation Oncology* Biology* Physics95, no. 2 (2016): 663-672.

[t] Nafees, B., Stafford, M., Gavriel, S., Bhalla, S. and Watkins, J., 2008. Health state utilities for non small cell lung cancer. Health and quality of life outcomes, 6(1), p.84.

Utilities	Utility score (SD)	Reference
HRQoL of Progressed Nafees 2008 (Progressive disease)	0.4734	Nafees 2008
QALY loss from surgery	0.011923077	Bendixen 2016[u]

Results

Stage I

Table 36. Stage I – Base case fully incremental results

Deterministic							
Cohort ID	Name	Absolute			Incremental		
					Fully incremental analysis		
		Costs	QALYs		Costs	QALYs	ICER
1	No Imaging	£985,211	29.88564				
2	CT then MRI	£1,114,291	32.89015		£129,079	3.00451	£42,962
3	MRI	£1,185,437	33.69798		£71,146	0.80783	£88,070

Table 37. Stage I – Base case results and scenario analyses

	ICER for CT followed by MRI vs No Imaging	ICER for MRI vs No Imaging	M ICER for RI vs CT followed by MRI
Base case	£42,962	£52,520	£88,070
PSA (5000 iterations)	£44,265	£52,127	£74,847

[u] Bendixen, M., Jørgensen, O.D., Kronborg, C., Andersen, C. and Licht, P.B., 2016. Postoperative pain and quality of life after lobectomy via video-assisted thoracoscopic surgery or anterolateral thoracotomy for early stage lung cancer: a randomised controlled trial. The Lancet Oncology, 17(6), pp.836-844.

Proportion 1-3 brain mets (Committee assumption)	£52,330	£57,702	£74,810
Brain mets detectable (71% - 5mm)	£49,744	£60,522	£100,605
Utility data for post progression survival and PFS (Nafees 2008)	£49,524	£60,527	£101,398
Survival curve – Brown	£46,275	£57,252	£99,012
Survival – TNM8	£50,382	£62,842	£110,098
Survival – Kocher	£42,129	£51,263	£84,772
Survival – GPA	£42,572	£51,992	£87,038
PFS – Brown (set extracranial progression to zero in decision tree as Brown data are only intracranial progression)	£42,156	£51,553	£86,458
Treatment with curative intent – all brain events	£36,191	£45,783	£81,458
Treatment with curative intent – intra-progression events only	£46,746	£56,286	£91,766
No acceleration factor to progression free survival curve for false negatives	£44,464	£54,319	£91,172
Acceleration factor for the Kocher progression free survival (FN 1-3 brain mets) (30%)	£39,445	£48,327	£80,943
Confirmatory MRI scan for all CT scanned patients	£43,565	£52,520	£85,827
Surgical temporary disutility removed	£46,590	£56,729	£93,746
Adenocarcinoma hazard ratio and adenocarcinoma prevalence	£28,574	£33,832	£51,867
Non-Adenocarcinoma hazard ratio and non- adenocarcinoma activity prevalence	£122,381	£163,575	£403,422
Sensitivity and specificity of MRI and CT from meta-analysis using the ‘mada’ package in R	£52,428	£57,348	£66,536
Confirmatory MRI scan for all CT scanned patients and MRI sensitivity at 0.6864	£50,333	£62,387	£110,703
Confirmatory MRI scan for all CT scanned patients and MRI sensitivity at 0.9991	£42,395	£50,845	£81,886
Confirmatory MRI scan for all CT scanned patients and CT sensitivity at 0.1154	£90,469	£52,520	£52,520 (CT-MRI extendedly dominated)

Confirmatory MRI scan for all CT scanned patients and CT sensitivity at 0.9971	£38,668	£52,520	£5,762,286

Table 38. Total strategy and strategy per patient cost for stage I patients

	CT Followed by MRI				MRI Only				No Imaging			
True status from model	**TP 1-3**	**TP 4+**	**FN 1-3**	**FN 4+**	**TP 1-3**	**TP 4+**	**FN 1-3**	**FN 4+**	**TP 1-3**	**TP 4+**	**FN 1-3**	**FN 4+**
Number of patients	19.96	10.04	8.49	0.00	26.8	10.0	1.7	0.0	0	0	28.45	10.04
Lobectomy	£14,783	£0	£31,577	£0	£19,814	£0	£6,299	£0	£0	£0	£105,859	£37,362
Wedge resection	£1,890	£0	£4,038	£0	£2,534	£0	£805	£0	£0	£0	£13,537	£4,778
Pneumonectomy	£1,143	£0	£2,441	£0	£1,531	£0	£487	£0	£0	£0	£8,182	£2,888
Complex and other resections	£362	£0	£773	£0	£485	£0	£154	£0	£0	£0	£2,591	£914
SABR	£39,318	£0	£6,860	£0	£52,698	£0	£1,368	£0	£0	£0	£22,998	£8,117
Standard Fractionated	£1,298	£0	£226	£0	£1,739	£0	£45	£0	£0	£0	£759	£268
Hypo Fractionated	£9,305	£0	£1,623	£0	£12,471	£0	£324	£0	£0	£0	£5,443	£1,921
CHART	£2,888	£0	£504	£0	£3,871	£0	£101	£0	£0	£0	£1,689	£596
SACT	£47,968	£24,127	£0	£0	£64,292	£24,127	£0	£0	£0	£0	£0	£0
Stereotactic radiosurgery	£53,240	£0	£0	£0	£71,357	£0	£0	£0	£0	£0	£0	£0
Surgical brain resection	£14,039	£0	£0	£0	£18,816	£0	£0	£0	£0	£0	£0	£0
WBRT	£2,172	£1,639	£0	£0	£2,911	£1,639	£0	£0	£0	£0	£0	£0
Radiotherapy for local control	£3,043	£14,159	£0	£0	£4,079	£14,159	£0	£0	£0	£0	£0	£0
Total for true status	£191,448	£39,924	£48,042	£0	£256,598	£39,924	£9,583	£0	£0	£0	£161,058	£56,844
Total for strategy	£279,415				£306,106				£217,902			
Number of people in strategy	38.49				38.49				38.49			
Cost per person within strategy	£7,259				£7,952				£5,661			

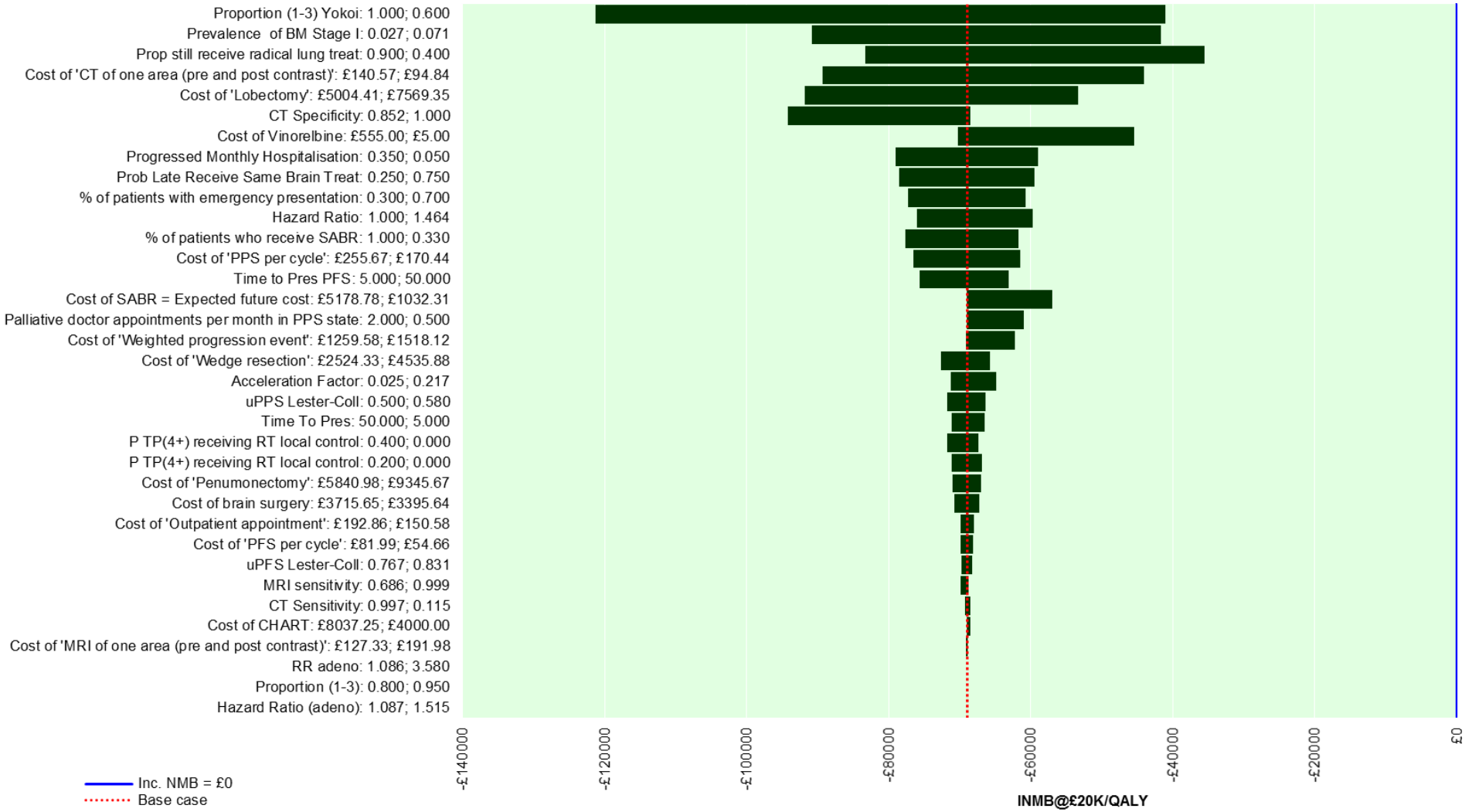


Figure 6. Stage I - CT then MRI vs No Imaging using INMB of £20,000/QALY (Base case ICER £42,962)

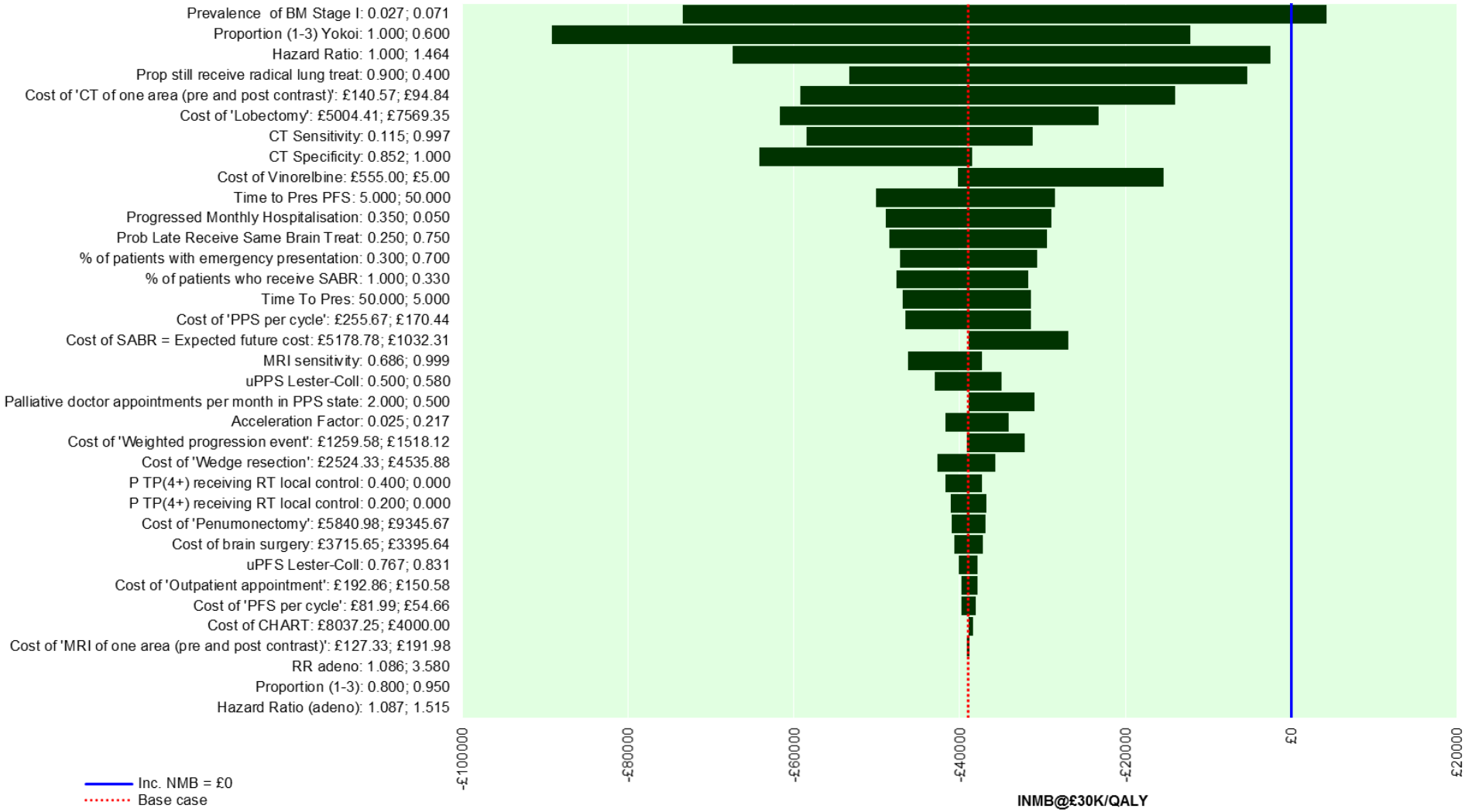


Figure 7. Stage I - CT then MRI vs No Imaging using INMB of £30,000/QALY (Base case ICER £42,962)

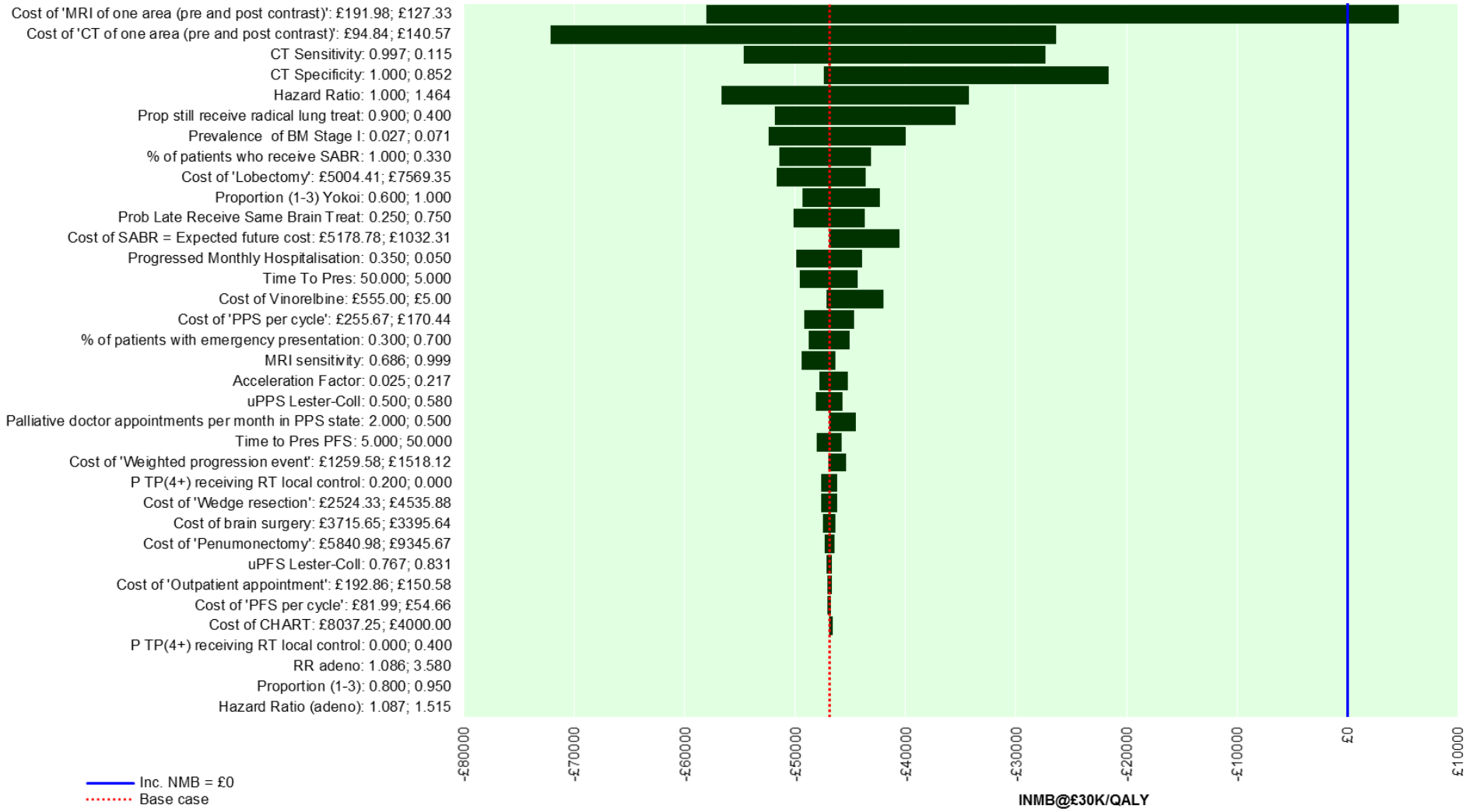


Figure 8. Stage I - MRI vs CT then MRI using INMB of £30,000/QALY (Base case ICER £88,070)

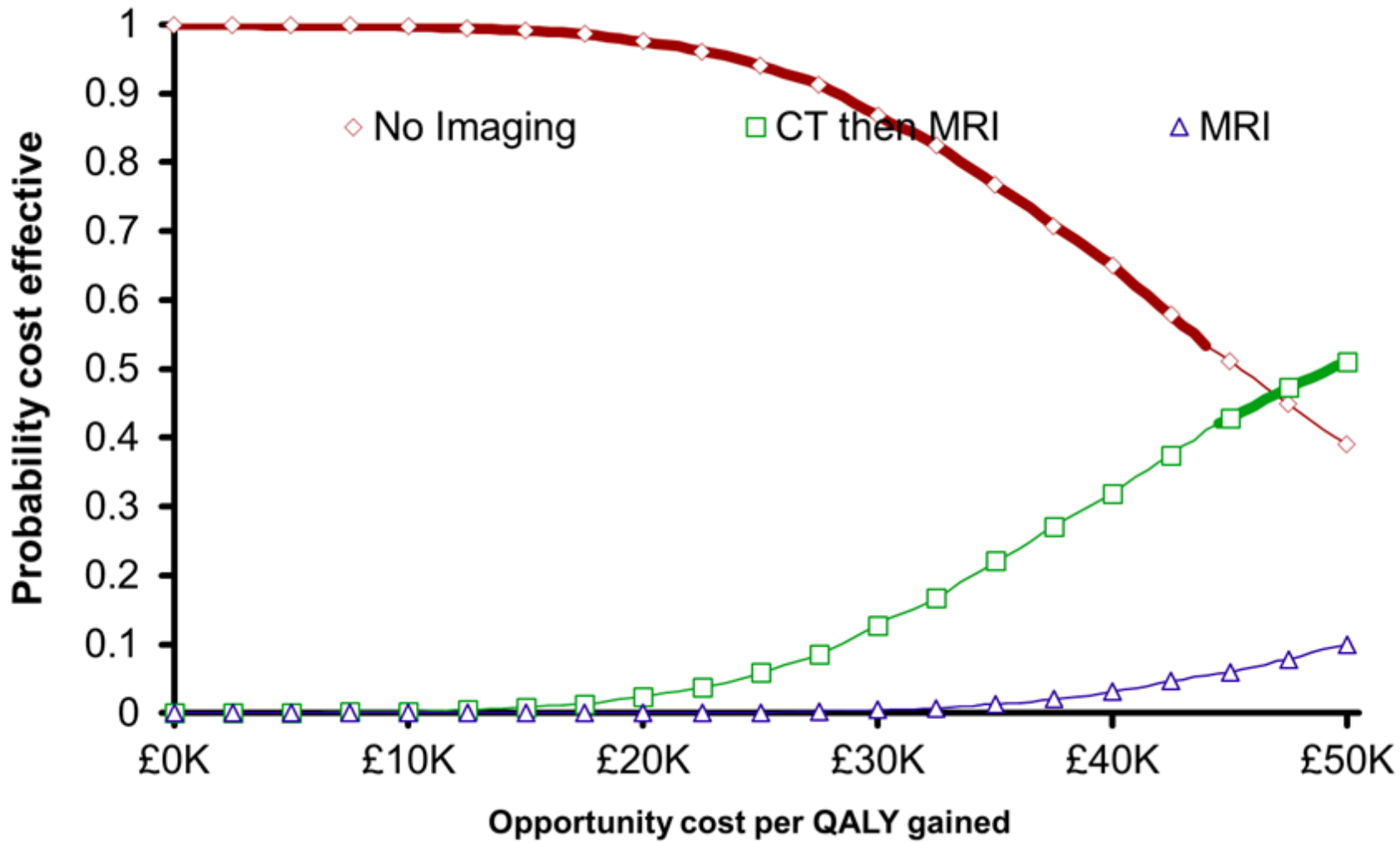


Figure 9. Stage I – Cost-effectiveness acceptability curve (CEAC) (5000 PSA iterations)

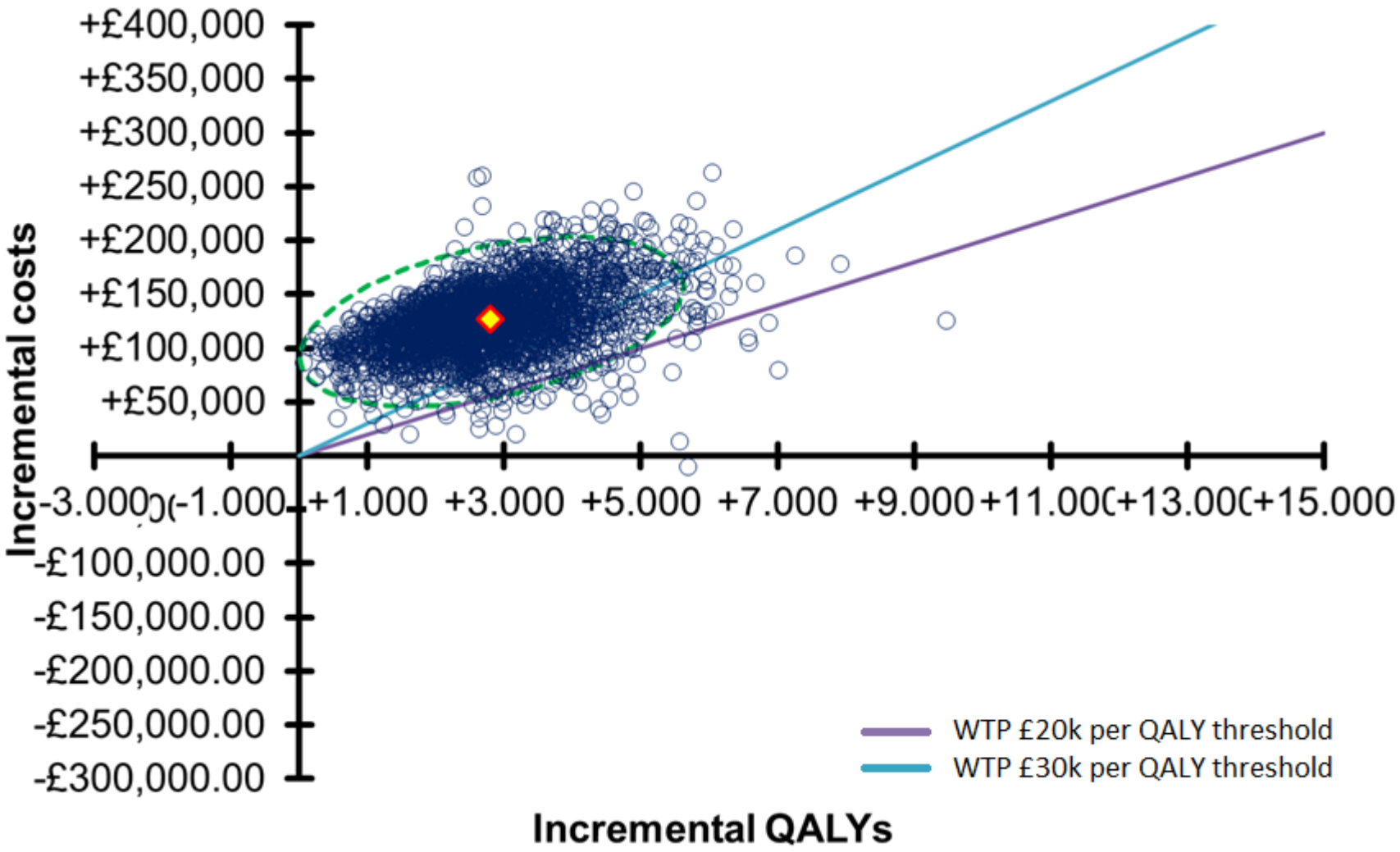


Figure 10. Stage I - CT followed by MRI compared to No Imaging (5000 PSA iterations)

Stage II

Table 39. Stage II – Base case fully incremental results

Cohort ID	Name	Absolute			Incremental		
					Fully incremental analysis		
		Costs	QALYs		Costs	QALYs	ICER
1	No Imaging	£2,006,903	61.37230				
2	CT then MRI	£2,137,057	67.54227		£130,153	6.16997	£21,095
3	MRI	£2,215,910	69.20121		£78,853	1.65894	£47,532

Table 40. Stage II – Base case results and scenario analyses

	ICER for CT followed by MRI vs No Imaging	ICER for MRI vs No Imaging	ICER for MRI vs CT followed by MRI
Base case	£21,095	£26,697	£47,532
PSA (5000 iterations)	£21,041	£26,256	£41,361
Proportion 1-3 brain mets (Committee assumption)	£30,536	£33,054	£41,073
Brain mets detectable (71% - 5mm)	£24,397	£30,593	£53,636
Utility data for post progression survival and PFS (Nafees 2008)	£24,317	£30,767	£54,725
Survival curve – Brown	£21,599	£27,974	£52,224
Survival – TNM8	£22,313	£29,557	£57,034
Survival – Kocher	£21,004	£26,389	£46,143
Survival – GPA	£21,024	£26,544	£210,879
PFS – Brown (set extracranial progression to zero in decision tree as Brown data are only intracranial progression)	£20,599	£26,102	£46,546

Treatment with curative intent – all brain events	£14,323	£19,959	£40,920
Treatment with curative intent – intra-progression events only	£24,879	£30,462	£51,227
No acceleration factor to progression free survival curve for false negatives	£22,144	£27,930	£49,565
Acceleration factor for the Kocher progression free survival (FN 1-3 brain mets) (30%)	£18,638	£23,823	£42,861
Confirmatory MRI scan for all CT scanned patients	£21,698	£26,697	£45,289
Surgical temporary disutility removed	£22,876	£28,836	£50,595
Adenocarcinoma hazard ratio and adenocarcinoma prevalence	£16,530	£19,854	£31,256
Non-Adenocarcinoma hazard ratio and non- adenocarcinoma activity prevalence	£49,139	£69,943	£191,078
Sensitivity and specificity of MRI and CT from meta-analysis using the ‘mada’ package in R	£22,843	£28,201	£38,207
Confirmatory MRI scan for all CT scanned patients and MRI sensitivity at 0.6864	£23,207	£29,771	£56,080
Confirmatory MRI scan for all CT scanned patients and MRI sensitivity at 0.9991	£21,437	£26,175	£43,579
Confirmatory MRI scan for all CT scanned patients and CT sensitivity at 0.1154	£29,634	£26,697	£26,697 (CT-MRI extendedly dominated)
Confirmatory MRI scan for all CT scanned patients and CT sensitivity at 0.9971	£20,869	£26,697	£2,428,676
% of stage II radical radiotherapy patients receiving SABR = 10%	£20,219	£25,646	£45,828

Table 41. Total strategy and strategy per patient cost for stage II patients

	CT Followed by MRI				MRI Only				No Imaging			
True status from model	**TP 1-3**	**TP 4+**	**FN 1-3**	**FN 4+**	**TP 1-3**	**TP 4+**	**FN 1-3**	**FN 4+**	**TP 1-3**	**TP 4+**	**FN 1-3**	**FN 4+**
Number of patients	41.00	20.62	17.43	0.00	55.0	20.6	3.5	0.0	0	0	58.43	20.62
Lobectomy	£30,358	£0	£64,846	£0	£40,689	£0	£12,935	£0	£0	£0	£217,390	£76,726
Wedge resection	£3,882	£0	£8,292	£0	£5,203	£0	£1,654	£0	£0	£0	£27,798	£9,811
Pneumonectomy	£2,346	£0	£5,012	£0	£3,145	£0	£1,000	£0	£0	£0	£16,803	£5,930
Complex and other resections	£743	£0	£1,587	£0	£996	£0	£317	£0	£0	£0	£5,320	£1,878
SABR	£40,371	£0	£7,044	£0	£54,109	£0	£1,405	£0	£0	£0	£23,614	£8,334
Standard Fractionated	£2,665	£0	£465	£0	£3,572	£0	£93	£0	£0	£0	£1,559	£550
Hypo Fractionated	£38,883	£0	£6,784	£0	£52,115	£0	£1,353	£0	£0	£0	£22,744	£8,027
CHART	£5,931	£0	£1,035	£0	£7,950	£0	£206	£0	£0	£0	£3,469	£1,224
SACT	£98,506	£49,546	£0	£0	£132,028	£49,546	£0	£0	£0	£0	£0	£0
Stereotactic radiosurgery	£109,332	£0	£0	£0	£146,537	£0	£0	£0	£0	£0	£0	£0
Surgical brain resection	£28,830	£0	£0	£0	£38,641	£0	£0	£0	£0	£0	£0	£0
WBRT	£4,460	£3,365	£0	£0	£5,978	£3,365	£0	£0	£0	£0	£0	£0
Radiotherapy for local control	£6,250	£29,076	£0	£0	£8,376	£29,076	£0	£0	£0	£0	£0	£0
Total for true status	£372,557	£81,988	£95,065	£0	£499,339	£81,988	£18,962	£0	£0	£0	£318,697	£112,481
Total for strategy	£549,610				£600,289				£431,178			
Number of people in strategy	79.05				79.05				79.05			
Cost per person within strategy	£6,953				£7,594				£5,455			

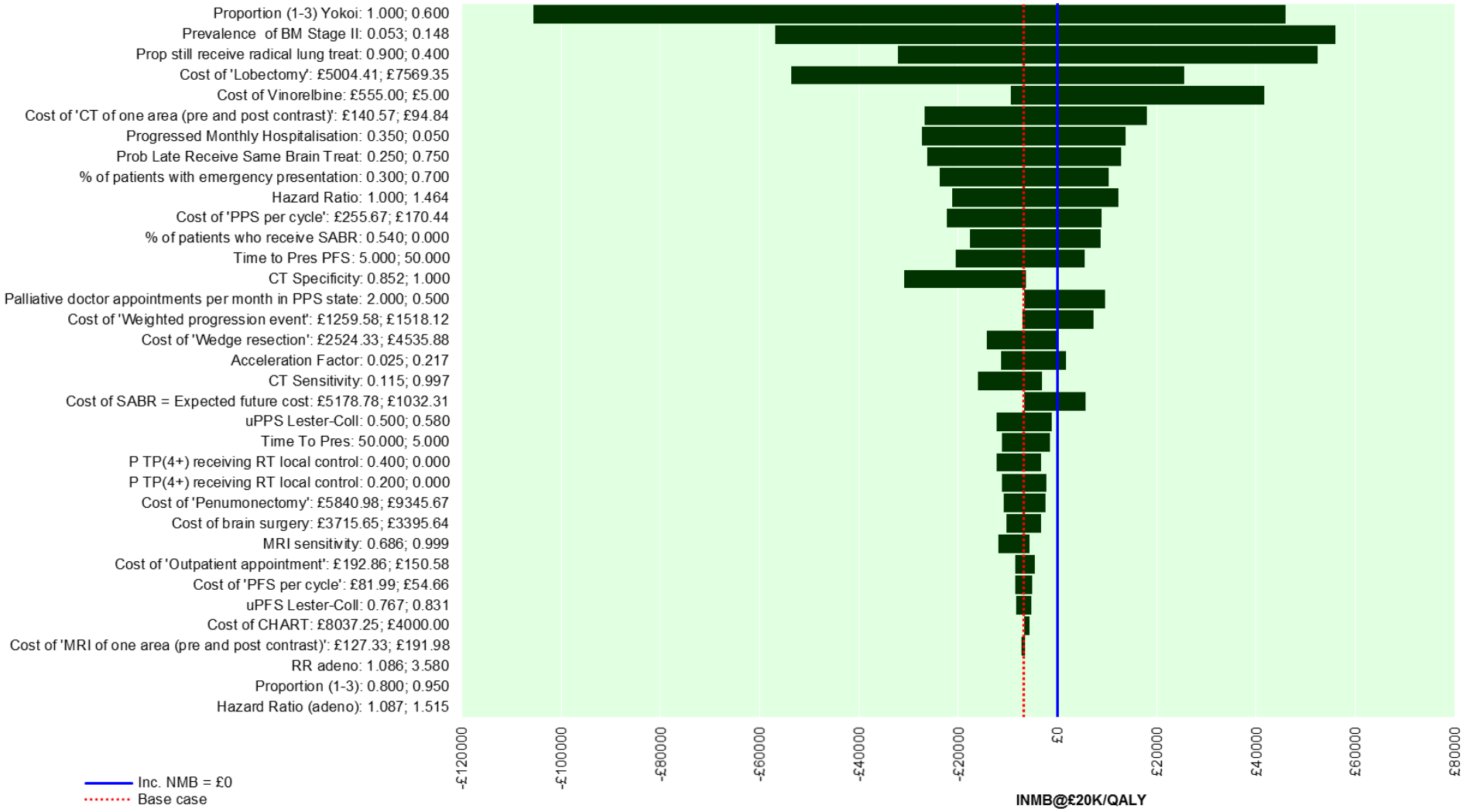


Figure 11. Stage II - CT then MRI vs No Imaging using INMB of £20,000/QALY (Base case ICER £21,095)

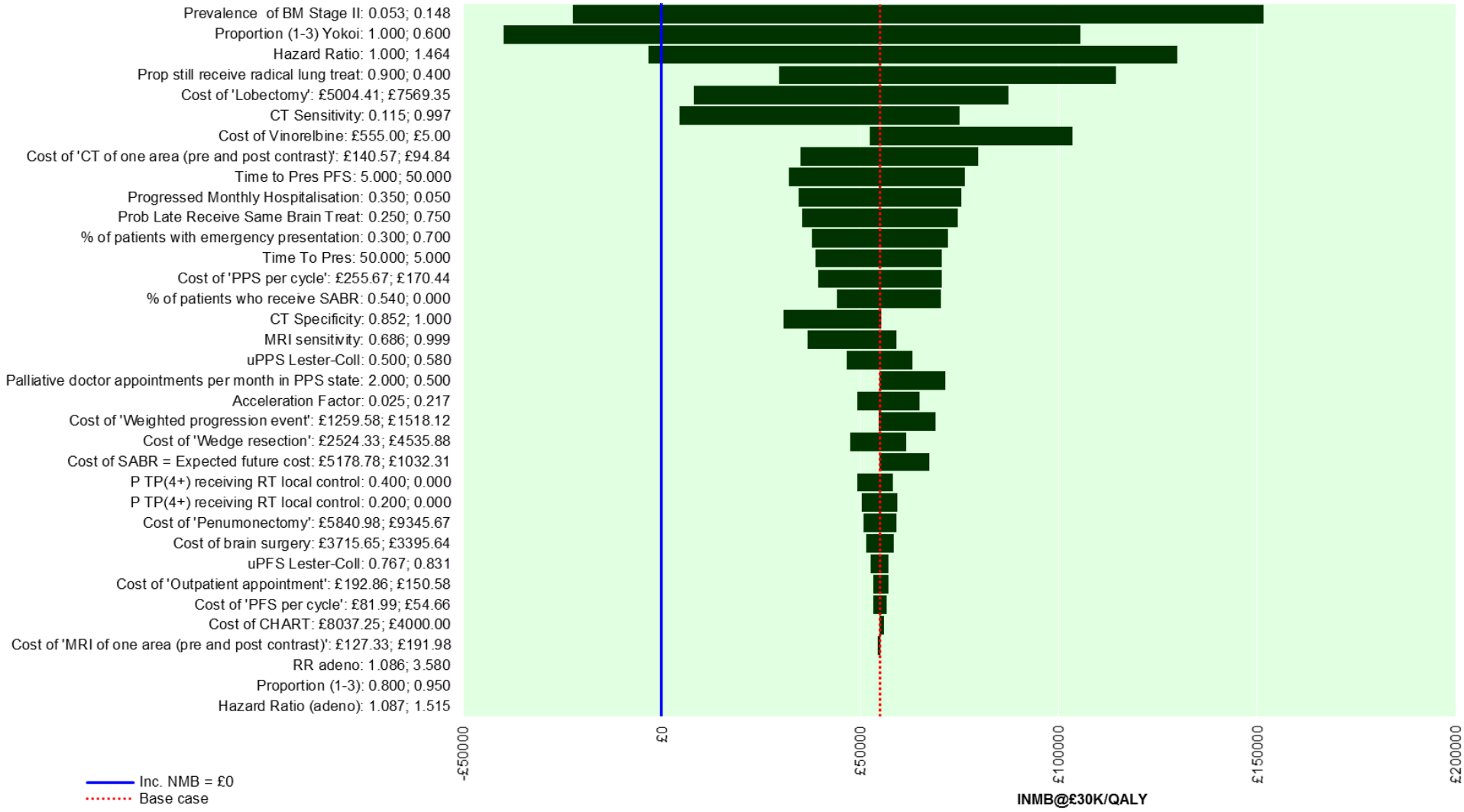


Figure 12. Stage II - CT then MRI vs No Imaging using INMB of £30,000/QALY (Base case ICER £21,095)

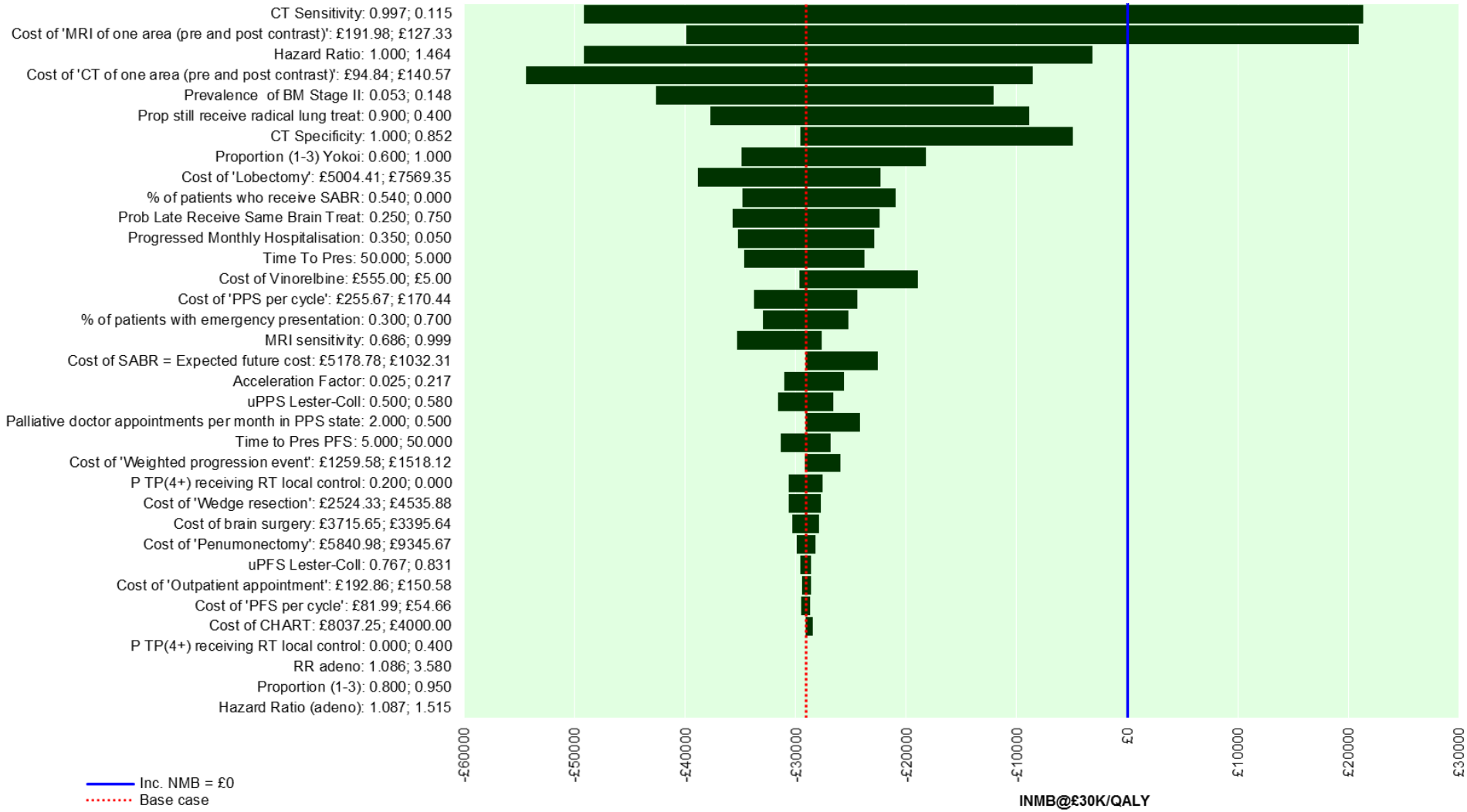


Figure 13. Stage II - MRI vs CT then MRI using INMB of £30,000/QALY (Base case ICER £47,532)

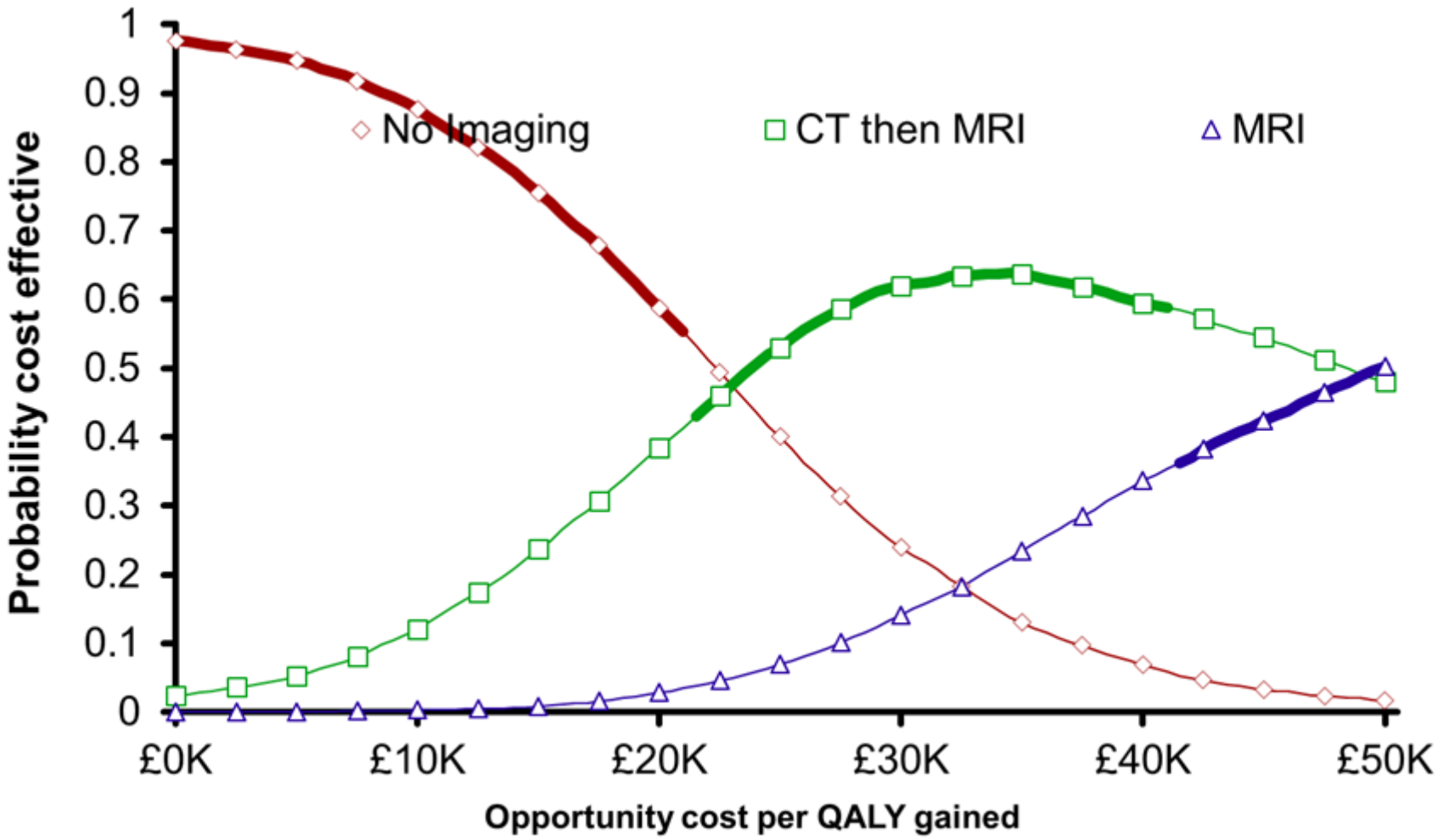


Figure 14. Stage II – CEAC (5000 PSA iterations)

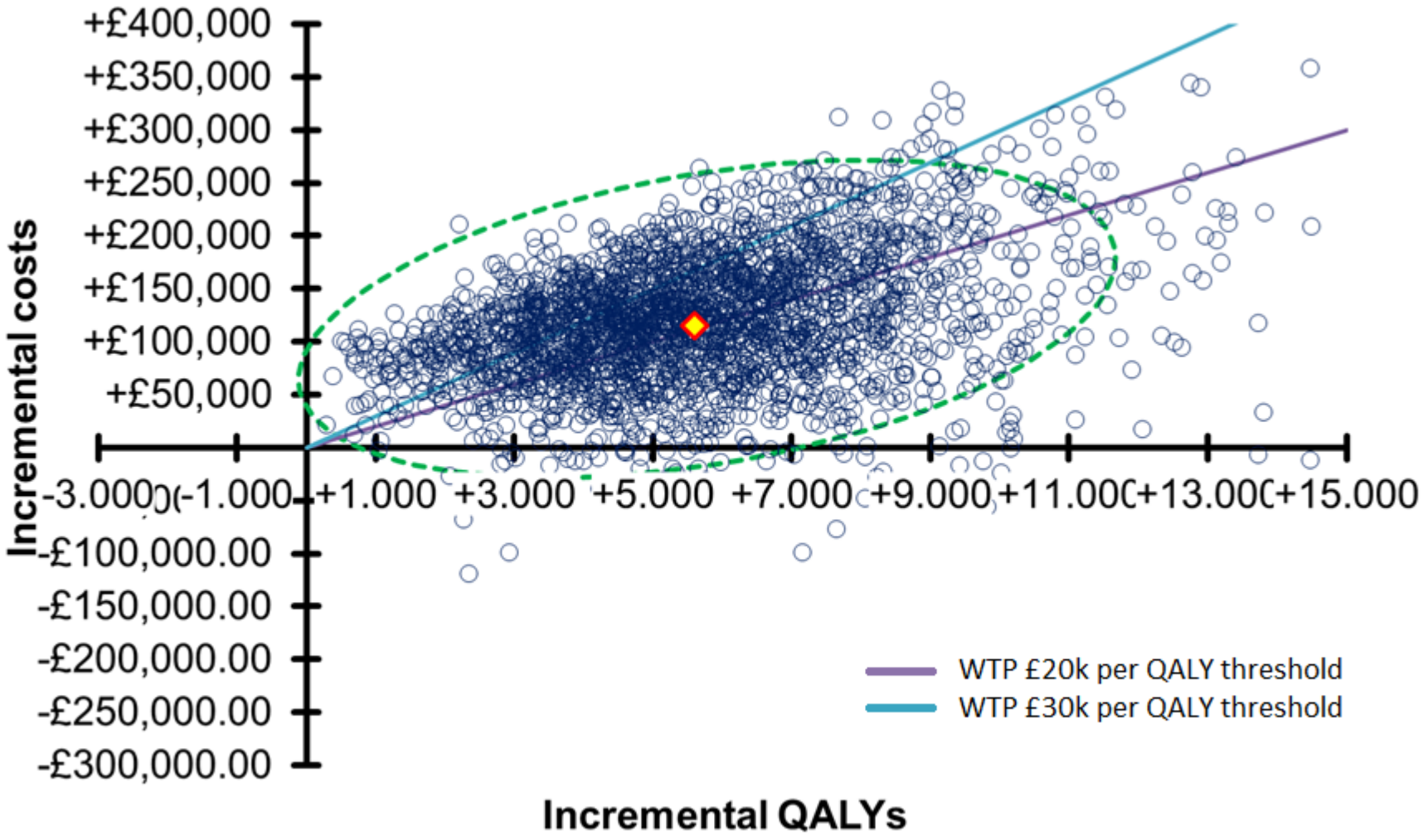


Figure 15. Stage II - CT followed by MRI compared to No Imaging (5000 PSA iterations)

Stage IIIA

Table 42. Stage III – Base case fully incremental results

Cohort ID	Name	Absolute			Incremental		
					Fully incremental analysis		
		Costs	QALYs		Costs	QALYs	ICER
3	MRI	£1,906,634	68.01546				
2	CT then MRI	£1,908,749	66.40836		£2,115	-1.60710	dominated
1	No Imaging	£2,032,817	60.48191		£126,183	-7.53355	dominated

Table 43. Stage IIIA - Base case results and scenario analyses

	ICER for CT followed by MRI vs No Imaging	ICER for MRI vs No Imaging	ICER for MRI vs CT followed by MRI
Base case	dominant[v]	dominant	dominant
PSA (5000 iterations)	dominant	dominant	dominant
Proportion 1-3 brain mets (Committee assumption)	dominant	dominant	dominant
Brain mets detectable (71% - 5mm)	dominant	dominant	£5,750
Utility data for post progression survival and PFS (Lester-Coll 2016)	dominant	dominant	dominant
Survival curve – Brown	dominant	dominant	dominant
Survival – TNM8	dominant	dominant	dominant
Survival – Kocher	dominant	dominant	dominant
Survival – GPA	dominant	dominant	dominant

[v] Dominant here refers to the intervention being less expensive and more effective than the comparator.

PFS – Brown (set extracranial progression to zero in decision tree as Brown data are only intracranial progression)	dominant	dominant	dominant
Curative intent – all brain events	dominant	dominant	dominant
Curative intent – intra-progression events only	dominant	dominant	£1,842
No acceleration factor to progression free survival curve for false negatives	dominant	dominant	dominant
Acceleration factor for the Kocher progression free survival (FN 1-3 brain mets) (30%)	dominant	dominant	dominant
Confirmatory MRI scan for all CT scanned patients	dominant	dominant	dominant
Surgical temporary disutility removed	dominant	dominant	dominant
Adenocarcinoma hazard ratio and adenocarcinoma activity prevalence	dominant	dominant	dominant
Non-Adenocarcinoma hazard ratio and non- adenocarcinoma activity prevalence	dominant	dominant	£22,824
Sensitivity and specificity of MRI and CT from meta-analysis using the ‘mada’ package in R	dominant	dominant	dominant
Confirmatory MRI scan for all CT scanned patients and MRI sensitivity at 0.6864	dominant	dominant	£8,768
Confirmatory MRI scan for all CT scanned patients and MRI sensitivity at 0.9991	dominant	dominant	dominant
Confirmatory MRI scan for all CT scanned patients and CT sensitivity at 0.1154	£4,839	dominant	dominant
Confirmatory MRI scan for all CT scanned patients and CT sensitivity at 0.9971	dominant	dominant	£2,850,523

Table 44. Total strategy and strategy per patient cost for stage IIIA patients

	CT Followed by MRI				MRI Only				No Imaging			
True status from model	**TP 1-3**	**TP 4+**	**FN 1-3**	**FN 4+**	**TP 1-3**	**TP 4+**	**FN 1-3**	**FN 4+**	**TP 1-3**	**TP 4+**	**FN 1-3**	**FN 4+**
Number of patients	40.23	20.24	17.10	0.00	53.9	20.2	3.4	0.0	0.0	0.0	57.33	20.24
Lobectomy	£0	£0	£49,793	£0	£0	£0	£9,932	£0	£0	£0	£166,927	£58,916
Wedge resection	£0	£0	£0	£0	£0	£0	£0	£0	£0	£0	£0	£0
Pneumonectomy	£0	£0	£3,849	£0	£0	£0	£768	£0	£0	£0	£12,902	£4,554
Complex and other resections	£0	£0	£0	£0	£0	£0	£0	£0	£0	£0	£0	£0
SABR	£0	£0	£0	£0	£0	£0	£0	£0	£0	£0	£0	£0
Standard Fractionated	£0	£0	£1,450	£0	£0	£0	£289	£0	£0	£0	£7,218	£2,547
Hypo Fractionated	£0	£0	£21,156	£0	£0	£0	£4,220	£0	£0	£0	£67,614	£23,864
CHART	£0	£0	£3,227	£0	£0	£0	£644	£0	£0	£0	£16,063	£5,669
SACT	£120,831	£60,775	£31,063	£0	£161,950	£60,775	£6,196	£0	£0	£0	£104,134	£36,753
Stereotactic radiosurgery	£14,305	£0	£0	£0	£19,173	£0	£0	£0	£0	£0	£0	£0
Surgical brain resection	£0	£0	£0	£0	£0	£0	£0	£0	£0	£0	£0	£0
WBRT	£17,509	£3,302	£0	£0	£23,467	£3,302	£0	£0	£0	£0	£0	£0
Radiotherapy for local control	£0	£28,533	£0	£0	£0	£28,533	£0	£0	£0	£0	£0	£0
Total for true status	£152,645	£92,610	£110,538	£0	£204,590	£92,610	£22,049	£0	£0	£0	£374,858	£132,303
Total for strategy	£355,793				£319,249				£507,161			
Number of people in strategy	77.57				77.57				77.57			
Cost per person within strategy	£4,587				£4,116				£6,538			

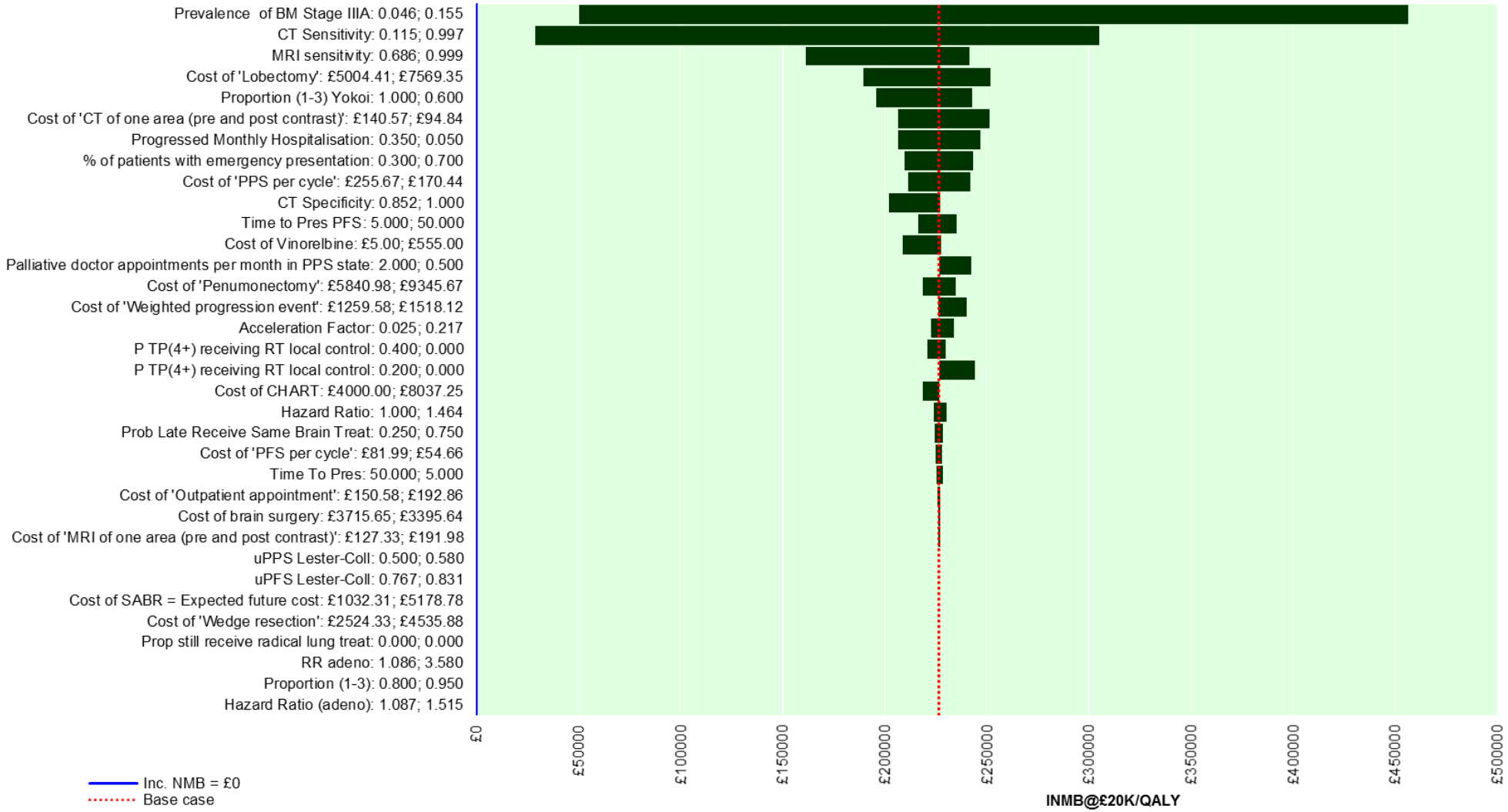


Figure 16. Stage IIIA - CT then MRI vs No Imaging using INMB of £20,000/QALY (Base case ICER situation was dominant)

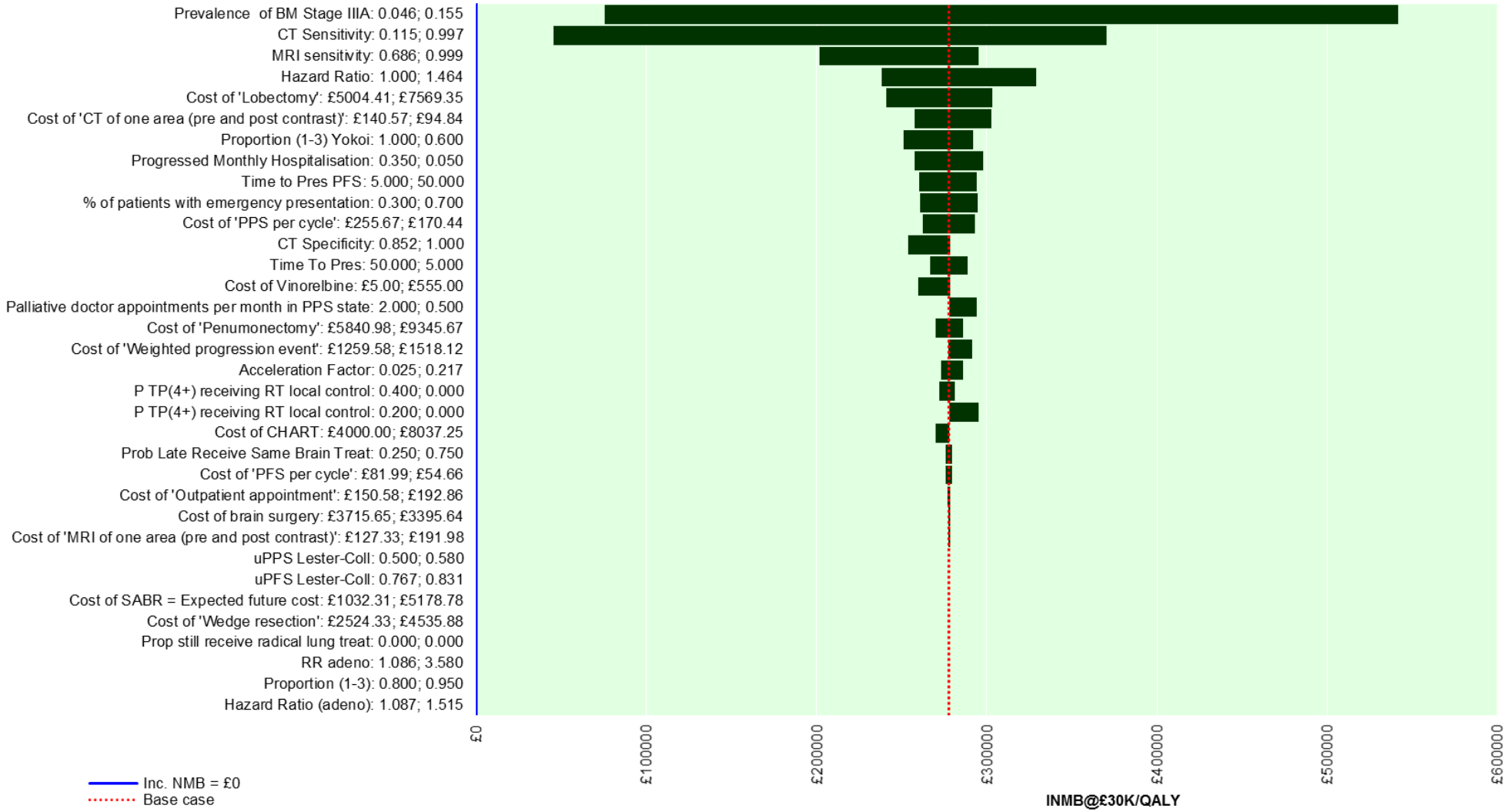


Figure 17. Stage IIIA - CT then MRI vs No Imaging using INMB of £30,000/QALY (CT-MRI is dominant in the base case)

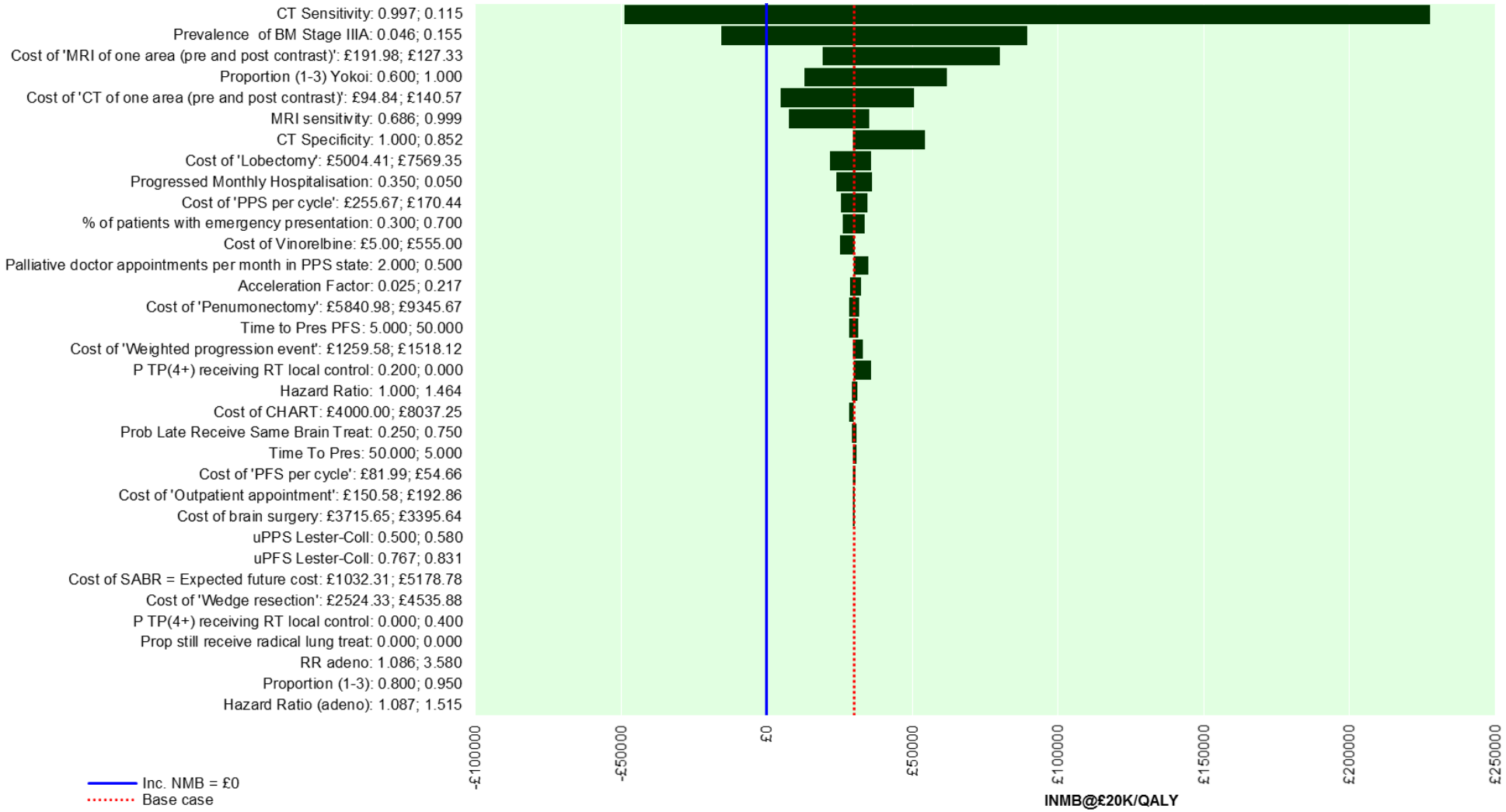


Figure 18. Stage IIIA - MRI vs CT then MRI using INMB of £20,000/QALY (MRI is dominant in the base case)

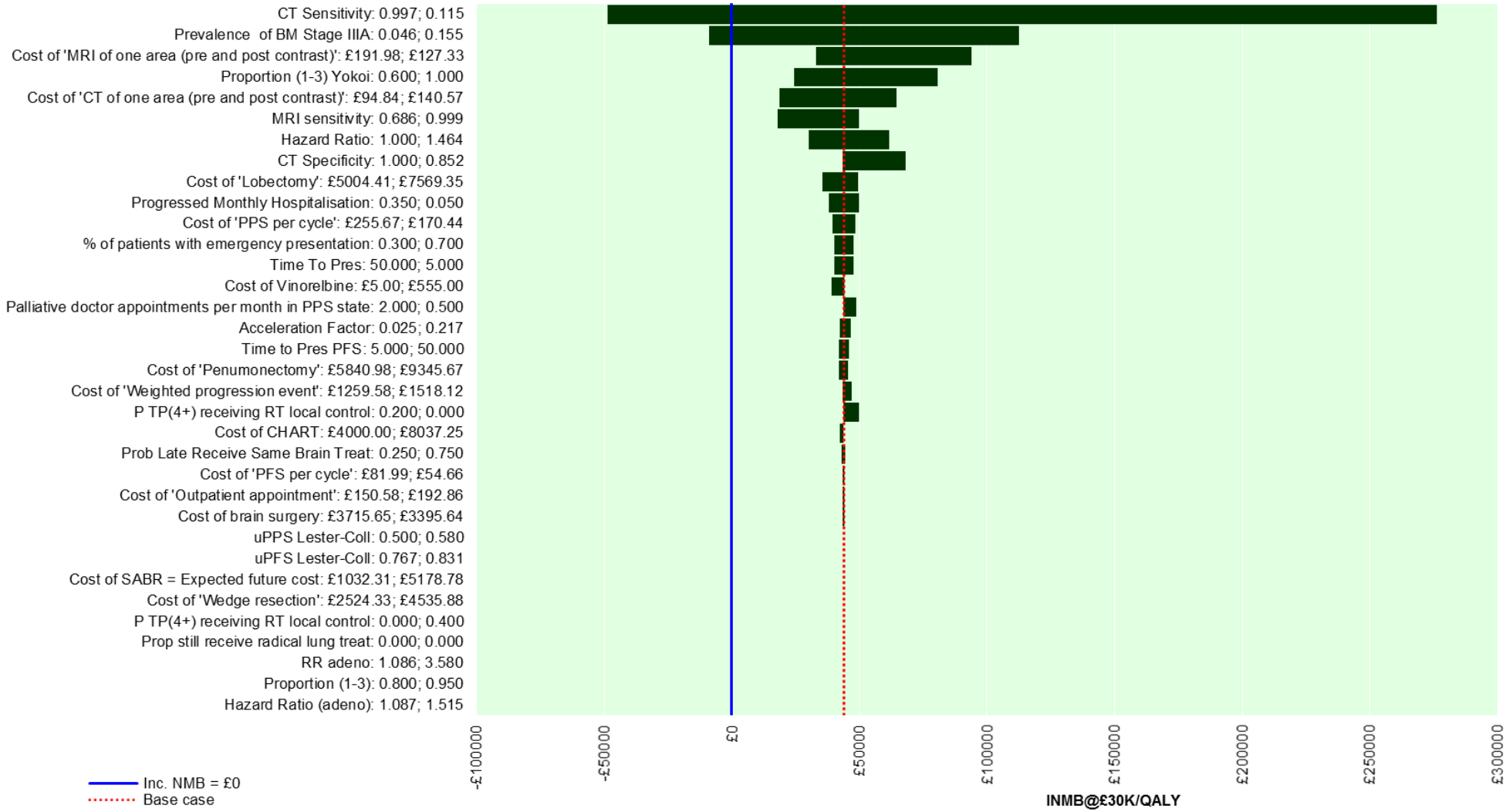


Figure 19. Stage IIIA - MRI vs CT then MRI using INMB of £30,000/QALY (MRI is dominant in the base case)

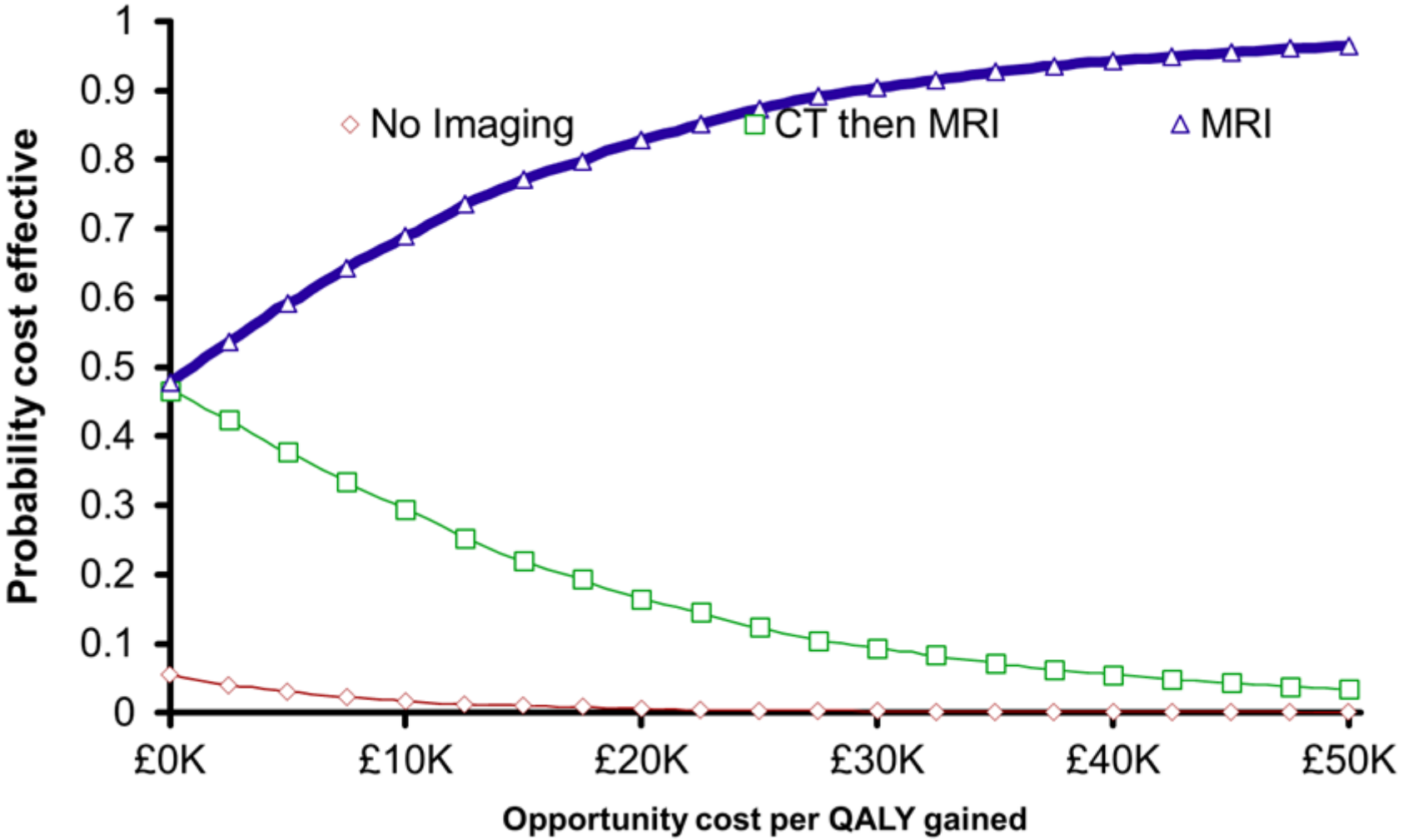


Figure 20. Stage IIIA – CEAC (5000 PSA iterations)

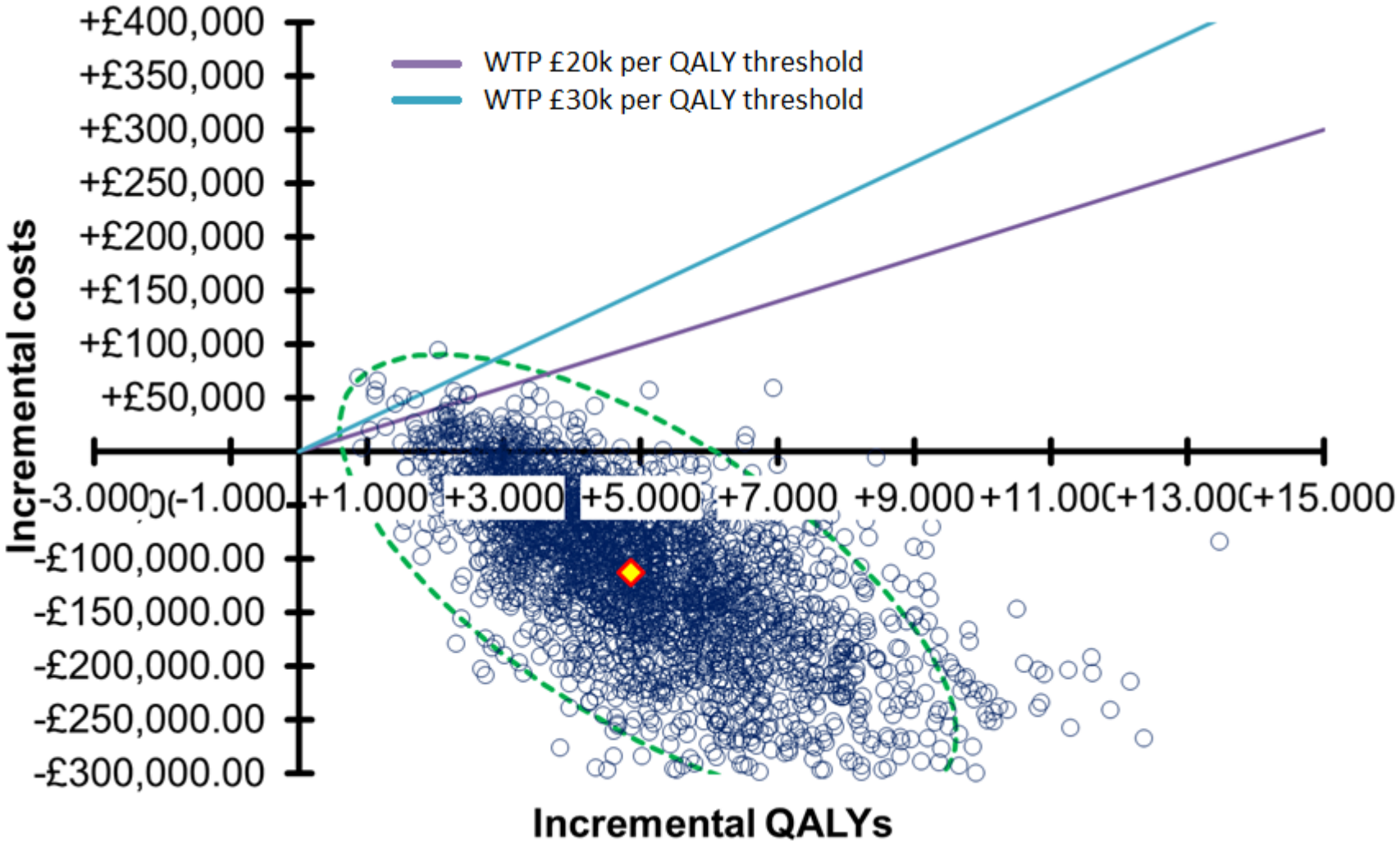


Figure 21. Stage IIIA - CT followed by MRI compared to No Imaging (5000 PSA iterations)

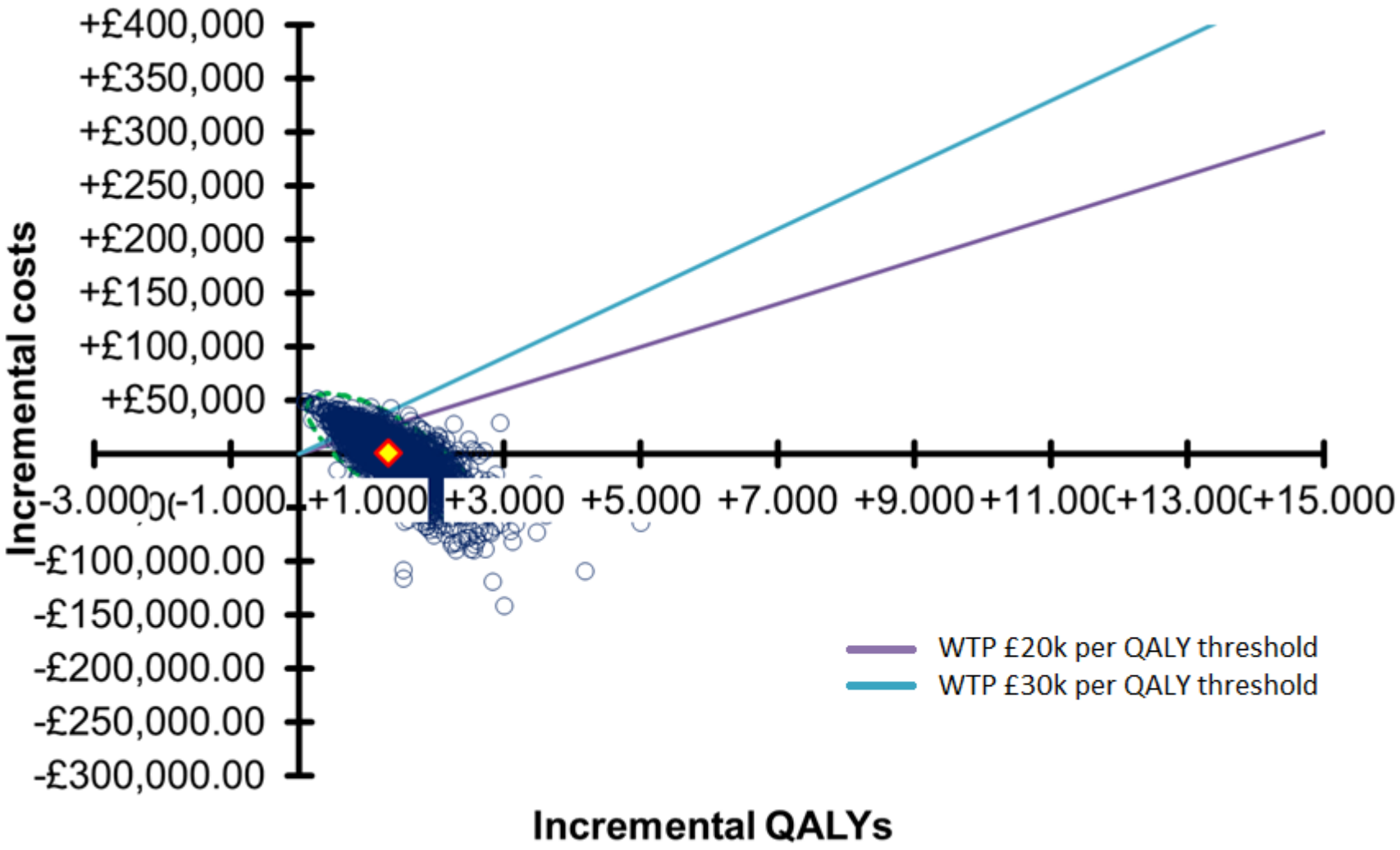


Figure 22. Stage IIIA – MRI compared to CT followed by MRI (5000 PSA iterations)

Discussion

This model calculated the number of cases of brain metastases (BM) that might be detected using MRI brain, CT brain followed by MRI brain, and no imaging strategies. The model combined the prevalence of brain metastases and the proportion detectable (as shown in Table 3) with the sensitivity of the test to calculate the number of true positive (1-3 or 4+), and false negative patients (1-3 and 4+) for each of the imaging strategies by NSCLC stage (Table 38, Table 41, Table 44). For stage I patients, MRI scanning alone produced 36.8 true positive and 1.7 false negatives per 1,000 patients imaged compared to 30.0 true positive and 8.49 false negatives for CT followed by MRI. For both strategies, 10 of the true positives have 4+ brain metastases and none of the false negatives do. For stage II patients, MRI scanning alone produced 75.6 true positive and 3.5 false negatives compared to 61.62 true positive and 17.43 false negatives for CT followed by MRI. For both strategies, 20 of the true positives have 4+ brain metastases and none of the false negatives do. For stage IIIA patients, MRI scanning alone produced 74.1 true positive and 3.4 false negatives compared to 60.47 true positive and 17.1 false negatives for CT followed by MRI. For both strategies, 20 of the true positives have 4+ brain metastases and none of the false negatives do.

If opportunity cost were not a concern, then it would be logical to give all patients who have received initial staging for their lung cancer and are being considered for radical treatment with curative intent an initially more expensive MRI scan (£180) because it is the most sensitive and jointly most specific strategy.. As the opportunity costs are important, the purpose of this economic analysis was to establish cost-effectiveness of these strategies at thresholds of £20,000 and £30,000 per QALY gained.

The key driving factors in this model was the overall prevalence of brain metastases, the proportion of positive patients with 4+ metastases and the costs of radical treatments. The prevalence of brain metastases used in this analysis (shown in Table 3) in stage II and III were similar to each other, both being around double that in stage I.

Base case, probabilistic sensitivity analysis, and sensitivity analyses showing the overall cost-effectiveness of the imaging strategies versus one another for all stages of NSCLC considered are presented in this report.

Stage I NSCLC

For stage I patients with NSCLC, the results table (Table 36) showed that all ICERS were above £30,000 per QALY, except for when an adenocarcinoma hazard ratio and prevalence were used. The one-way sensitivity analysis (OSA) of CT followed by MRI compared to No Imaging when QALYs are valued at £20,000 (Figure 6) showed that no plausible variations in any of the parameters could make CT followed by MRI cost-effective compared to No Imaging. However, for the same analysis, when QALYs are valued at £30,000 (Figure 7), the upper bound of the 95% confidence interval for the prevalence of brain metastases could make CT followed by MRI cost effective compared to no imaging. The OSA of MRI compared to CT followed by MRI when QALYs are valued at £30,000 (Figure 8) showed that the only situation where MRI could be cost-effective compared to CT followed by MRI was when the cost of MRI scanning of one area (with pre and post contrast) was at its lowest possible value of £127.33.

For stage I patients with NSCLC, the results of the probabilistic sensitivity analysis were very similar to the base case results. The cost-effectiveness acceptability curve (CEAC) (Figure 9) showed that we would have to be prepared to pay around £46k/QALY for the probability of cost-effectiveness of CT followed by MRI to be as high as no imaging. On the graph of the PSA of 5000 iterations of CT followed by MRI compared to 'no imaging' (Figure 10), we can see that the average iteration marker (yellow diamond with the red border) is firmly above the light blue line denoting a threshold of £30,000/QALY. The majority of the density of the 5,000 iterations are above the above the £30,000 per QALY threshold line.

Based on these results, we can conclude that no imaging strategy involving the use of either technology (CT or MRI) for detecting brain metastases in stage I NSCLC patients prior to radical treatment with curative intent is cost-effective at willingness-to-pay thresholds of £20,000 or £30,000 per QALY. This is primarily due to the low prevalence of detectable brain metastases in the stage I population (~3.8%). Varying this value to the highest extreme of its confidence interval yielded an ICER of £29,067 per QALY for CT followed by MRI compared to no imaging.

Stage II NSCLC

For patients with stage II NSCLC, we carried out the same analysis as we carried out for stage I NSCLC patients. The only difference was the prevalence of detectable BM (~8%). In the deterministic base case, we found that ICER for CT followed by MRI was £21,095 – just over the threshold of £20,000 per QALY gained, but well under £30,000 per QALY. The ICER for MRI alone compared to CT followed by MRI was well in excess of £30,000 per QALY.

The results of the PSA followed a very similar pattern to the deterministic base case. The scatterplot of 5,000 PSA iterations (Figure 15) shows the average iteration marker between the dark purple line denoting a threshold of £20,000 per QALY, and the light blue line denoting a threshold of £30,000 per QALY. Most of the iterations fall evenly on either side of both of these lines demarcating these thresholds, showing reasonable uncertainly in the average ICER in relation to the common decision thresholds.

Of the 22 scenario analyses we performed shown in Table 38, only two scenarios (where the proportion of patients with 1-3 brain metastases came from the committee, and where non-adenocarcinoma hazard ratios and prevalence were used) exceeded the threshold of £30,000 per QALY. Three of these scenario analyses (where treatment with curative intent for all brain events, 'Acceleration factor for the Kocher progression free survival (FN 1-3 brain mets) (30%)' and 'Adenocarcinoma hazard ratio and adenocarcinoma prevalence') ICERs were below the £20,000 per QALY threshold.

For stage II patients with NSCLC, neither for the base case, the PSA or any of the incremental analysis for MRI alone when compared to CT followed by MRI shown in Table 40 had an ICER below the £30,000 per QALY threshold.

The OSA for CT followed by MRI compared to 'no imaging' at a willingness-to-pay threshold of £20,000 per QALY (Figure 11) showed that model was sensitive to a large number of parameters, which when varied within their plausible ranged could cause the INMB to be above zero, therefore rendering CT followed by MRI a cost-effective strategy. A further OSA analysis of the same pairwise comparison, using a willingness-to-pay threshold of £30,000 per QALY (Figure 12), showed that only three parameters (prevalence of brain metastases in stage II, proportion of patients with 1-3 brain mets (Yokoi), and the hazard ratio), would be able to take the INMB into negative territory, thus rendering CT followed by MRI not cost-effective in comparison to the 'no imaging strategy'. The final OSA conducted for stage II NSCLC (Figure 13) showed that just two parameters (sensitivity of CT and the cost of MRI of one area) when varied within their plausible range, could render MRI cost-effective as compared to CT followed by MRI, a willingness-to-pay threshold of £30,000 per QALY.

The CEAC for stage II patients with NSCLC (Figure 14) shows that 'no imaging' strategy has the highest probability of being cost-effective until around a willingness-to-pay of £23,000 QALY, at which point it is equally likely to be as cost-effective as CT followed by MRI at 48%. From here, as the willingness-to-pay increases, the probability of CT followed of MRI being the most cost-effective strategy increases until around a willingness-to-pay threshold of £34,000 per QALY where the probability is around 67%. At a willingness-to-pay threshold of £30,000 per QALY, CT followed by MRI has the highest probability of being the most cost-effective strategy at around 62%, whilst the 'no imaging' strategy has a probability of around 22%.

We can conclude with a fair amount of certainty that CT followed by the MRI is the most cost-effective strategy at a willingness-to-pay threshold of £30,000 per QALY for detecting brain metastases in stage II NSCLC patients prior to radical treatment with curative intent but it is uncertain whether the 'true' ICER for imaging lies above or below the £20,000 threshold.

Stage IIIA NSCLC

As discussed previously, the stage IIIA NSCLC patient group consist of five broad treatment strategies; those treated with Chemotherapy and Surgery (CS), Chemotherapy and Radiotherapy (CR), Chemotherapy, Radiotherapy and Surgery (CRS), Radiotherapy only (R) and Surgery only (S). We combined and weighted each of the treatment strategies for stage IIIA patients into a single treatment strategy within the model, with the split between each of the treatment shown in Table 23.

In the base case, the PSA with 5,000 iterations and every analysis shown in Table 43, CT followed by MRI is a dominant strategy as compared to the no imaging strategy (which means that CT followed by MRI produced more benefits and cost less as compared to the no imaging strategy). In the base case, PSA and 16 of the 21 sensitivity analyses presented in the same table, MRI compared to CT followed by MRI, was a dominant strategy (meaning that MRI produced more benefits and cost less than CT followed by MRI). Of the strategies where MRI only was not dominant compared to CT followed by MRI, three had ICERS below £20,000 per QALY, one had an ICER between £20,000 and £30,000 per QALY, and one had an ICER above £30,000 per QALY.

The OSA associated with the stage IIIA analysis of CT followed by MRI compared to no imaging when QALYs are worth £20,000 (Figure 16) and £30,000 (Figure 17) both show that no parameter varied within their plausible threshold was able to make CT followed by MRI cost-ineffective. A further two analyses of MRI only compared to CT followed by MRI where QALYs are worth £20,000 (Figure 18) and £30,000 (Figure 19) show that the parameters concerned with the sensitivity of CT scanning, when increased to 0.997, and the parameter concerned with the prevalence of brain metastases in stage IIIA patients, when lowered to 0.046, could render MRI cost-ineffective at both these thresholds.

The CEAC for stage IIIA (Figure 20) shows that the MRI only strategy has an equal chance of being cost-effective compared to CT followed by MRI at a willingness-to-pay threshold of £0 per QALY, whilst the probability of the 'no imaging' strategy is around 6%. As the willingness-to-pay increases, the probability of the MRI only strategy being the most cost-effective also increase whilst both CT followed by MRI and no imaging decrease. In Figure 21 showing 5,000 PSA iterations of CT followed by MRI compared to 'no imaging', we can see that the average iteration marker is firmly in the south east quadrant, showing that the average cost of the of CT followed by MRI as compared to the 'no imaging' strategy was lower, and produced more QALYS, and thus rendering CT followed by MRI a dominant strategy for this comparison. Furthermore, the vast majority of the iterations on this figure fall below the dark purple line demarcating a threshold of £20,000 per QALY, which in turn gives us considerable confidence that the ICER is below £20,000 per QALY.

A further similar comparison of MRI alone compared to the CT followed by MRI strategy (Figure 22) showed that the average iteration marker is still in the south-east quadrant, meaning that MRI alone is a dominant strategy as compared to CT followed by MRI, although not as pronounced as CT followed by MRI compared to 'no imaging.

Based on these results, we can conclude that for people with stage IIIA NSCLC, CT followed by MRI is preferable to the 'no imaging' strategy as it is dominant. However, a further pairwise comparison of MRI alone as the sole imaging strategy as compared to CT followed by MRI shows MRI alone to be the dominant strategy, and therefore the overall most cost-effective strategy for detecting brain metastases in stage IIIA NSCLC patients prior to radical treatment with curative intent.

In summary:-

- No imaging strategy was cost-effective for stage I patients, mainly because of the low prevalence of BM.
- CT-MRI could be considered cost effective compared to no imaging for stage II patients. MRI is not cost-effective compared to CT-MRI, mainly because CT has a good sensitivity for identifying patients who are TP (4+), who contribute the most cost-benefit in the model.
- MRI is cost effective for stage IIIA patients, mainly because it is the most sensitive test and identifying a case contributes both QALY gains and cost savings

Strengths and Limitations

Our model has a number of important strengths; it is the only directly applicable health economic model to examine whether NSCLC patients selected for curative intent should receive brain imaging in a UK setting and includes a number of original pieces of evidence synthesis for survival and diagnostic accuracy data. We made use of a wide range of sensitivity and scenario analyses to explore the uncertainty in the model and can be confident that our conclusions, certainly for stages I and IIIA, are robust to plausible variations in parameters.

The model is also characterised by a number of important limitations; the diagnostic accuracy data was of low quality, the prevalence data came from a retrospective analysis, the proportion of people with 4+ brain mets was an important but highly uncertain parameter, the costs of systemic therapy were crudely captured, the effectiveness of treatment pathways was crudely captured, the survival curves and progression data were drawn from partly indirect populations and a large amount of parameters were underpinned by committee assumptions (the proportion of patients receiving different potential treatments upon diagnosis and progression, the health state occupancy costs and the consequences upon presentation). We also had to make a number of assumptions about the way that survival curves for the different groups were related to each other and the way that False Negative patients' progression would be accelerated. While we think that all of these assumptions were justified and we tested them in sensitivity analysis, they are not based on directly observed data in the population of interest (although this limitation is common to at least some populations in all economic models examining diagnostic test accuracy).

Appendix J – Research recommendations

• Question	• What is the effectiveness and cost-effectiveness of performing contrast enhanced CT brain routinely at the time of initial diagnosis/staging CT in people with suspected lung cancer?
Population	All patients with suspected lung cancer
Characteristics of interest	Sensitivity Specificity Accuracy of diagnosis and staging Impact on diagnostic pathway
Study design	Randomised controlled trial

• Potential criterion	• Explanation
Importance to patients, service users or the population	All patients with suspected lung cancer receive an initial diagnostic and staging CT scan. Adding a contrast enhanced CT of the brain at this time represents a small opportunity cost, both to the NHS and the patient and may help to streamline the diagnostic pathway, clarifying at the earliest possible time whether patients are suitable for treatment with curative intent or not. Furthermore, treatment for brain metastases in advanced disease is becoming more common, and these people would not normally receive brain imaging until presenting with symptoms. Their outcomes might be improved with earlier diagnosis.
Relevance to NICE guidance	Medium priority: a recommendation was made on the use of contrast enhanced MRI brain if there is a suspicion of brain metastasis on CT in patients with stage II NSCLC being treated with curative intent. A further recommendation has been made to offer contrast enhanced MRI brain for patients with stage IIIA NSCLC being treated with curative intent. While it is a recommended imaging modality, the sensitivity and specificity of CT in

Lung cancer: diagnosis and management: Evidence review for the clinical and cost-effectiveness of routine MRI or CT of the brain in the management of people with lung cancer prior to radical therapy with curative intent (March 2019)

• Potential criterion	• Explanation
	this population are also somewhat uncertain and this research would help to resolve these uncertainties.
Current evidence base	The quality of the evidence included in the clinical review ranged from very low to moderate, therefore there is a need for higher quality RCT evidence.
Equality	This study could improve equality of access to brain imaging prior to treatment with curative intent. This could impact significantly on the treatment plan.
Feasibility	There is a large enough population of people with this condition and the brain imaging techniques are available in current clinical practice.

Appendix K – WinBUGS Code

This codeset was used to meta-analyse the diagnostic test accuracy data for use in the model. It includes data from all the studies included in the clinical review minus Yokoi 1999 because the committee wished to exclude it through lack of clinical plausibility. The example below uses data from the studies reporting sensitivity and specificity for MRI. The same codeset was used for the CT data.

```
Random Effects
model{

for(i in 1:4){
                N1[i] <- tp[i] + fn[i]  # Number of patients with disease
                tp[i] ~ dbin(tpr[i], N1[i])
                logit(tpr[i]) <- lt[i]
                lt[i] ~ dnorm(mean1, prec1)

                N0[i] <- tn[i] + fp[i]  # Number of patients without disease
                fp[i] ~ dbin(fpr[i], N0[i])
                logit(fpr[i]) <- lf[i]
                lf[i] ~ dnorm(mean0, prec0)

}

# Vague priors:
mean1 ~ dnorm(0, 0.01)  # Mean logit(tpr)
sd1 ~ dunif(0,5) # Between-study SD in logit(tpr)
mean0 ~ dnorm(0, 0.01)  # Mean logit(fpr)
sd0 ~ dunif(0,5) # Between-study SD in logit(fpr)

prec1 <- pow(sd1, -2) # Precision
prec0 <- pow(sd0, -2) # Precision
```

```
logit(summtpr) <- mean1 # Define summary TPR
logit(summfpr) <- mean0 # Define summary FPR
summspec <- 1 - summfpr # Summary specificity


}

# Initial values:
list(mean1 = 0, sd1 = 1, mean0 = -1, sd0 = 0.5)
list(mean1 = 2, sd1 = 0.5, mean0 = -2, sd0 = 1)

# Data:
tp[]	fn[]	fp[]	tn[]
6	0	0	23
5	0	0	51
10	1	0	86
37	6	7	392


END
```