

Evidence-based Practice Center Systematic Review Protocol

Project Title: Imaging for Pretreatment Staging of Small Cell Lung Cancer

I. Background and Objectives for the Systematic Review

Epidemiology

Lung cancer is the most common cause of cancer death, and in 2014 will account for about 27% of cancer deaths in the United States.¹ Small cell lung cancer (SCLC) comprises approximately 15% of all lung cancers.² The five-year survival rate for SCLC is only 6%, which is only one-third of the corresponding rate for non-small-cell lung cancer (18%).¹ One subtype of SCLC is combined SCLC, a tumor primarily comprised of SCLC, but also including some NSCLC tumor.

In 2011 (the most recent year for which United States prevalence data are available), an estimated 402,236 people in the United States were living with lung cancer.³ In 2014, approximately 224,210 new cases of lung cancer will be diagnosed,¹ and roughly 15% (30,000-35,000) will be SCLC. The majority of patients diagnosed with SCLC are current or former smokers.⁴

Diagnosis

In evaluating patients with possible lung cancer, key clinical inputs include symptom assessment, clinical exam, risk factor assessment (e.g., smoking), and imaging such as chest X-ray and chest computed tomography (CT). However, SCLC remains a pathological diagnosis requiring tissue obtained via biopsy. Thus, while imaging plays a key role in the diagnostic workup, these tests cannot be used to definitively distinguish SCLC from NSCLC, other types of lung cancer, or other non-cancerous conditions (i.e., differential diagnosis). Multidetector CT is widely used as the critical imaging test to determine whether, where, and how to biopsy. Biopsy may be performed via one of several methods (e.g., bronchoscopy, CT-guided percutaneous biopsy, endobronchial ultrasound (EBUS), endoscopic ultrasound (EUS)), depending on the location and size of the mass as well as patient factors.

Although most cases are diagnosed after patients become symptomatic with cough, shortness of breath, weight loss or hemoptysis, SCLC can also be detected incidentally. In these asymptomatic patients, a chest X-ray or chest CT ordered for other reasons can reveal a nodule or suspicious mass requiring further investigation. Once a suspicious lung nodule or mass is detected, these patients are diagnosed and staged in the same manner as symptomatic patients.

Staging

Staging involves determining the extent of disease and informs the choice of treatment. Small cell lung cancer is typically staged using one of two systems:

- The first system, the Veterans Administration Lung Study Group (VALSG) system,⁵ classifies SCLC as either “limited stage” or “extensive stage” disease with the following definitions:
 - Limited stage disease (LD): Cancer is confined to one hemithorax and may be present in the regional lymph nodes or in ipsilateral supraclavicular nodes, all of which can be encompassed in a safe radiotherapy field.
 - Extensive stage disease (ED): Cancer that cannot be classified as limited stage, such as when there is presence of contralateral hilar or supraclavicular nodes, malignant pericardial or pleural effusions, or distant metastatic disease.
- The second system, which is used less commonly for SCLC but often used for other types of cancer, is the American Joint Committee on Cancer (AJCC) TNM system, where T stands for primary tumor, N stands for regional lymph nodes, and M stands for distant metastasis. Lung cancers are classified based on the size of the main tumor, whether it has locally invaded other organs/tissues, spread to lymph nodes, and whether it has metastasized to other parts of the body. This information is used to assign a stage between I and IV. A higher stage represents a more extensive spread of the cancer.

The National Cancer Institute reported that from 1975-2008, about 70% of SCLC cases presented with extensive stage disease, another 21% had regional spread such as mediastinal nodal involvement, and only 5% were localized (the other 4% were unstaged).⁶ The most common sites of metastases for SCLC are liver, adrenal glands, bone, bone marrow, and brain.⁷

Some debate exists about which system should be used. The International Association for the Study of Lung Cancer (IASLC) recommended in 2007 that clinicians use the AJCC system.⁸ This recommendation was based on an analysis of over 8,000 patients with SCLC, which found that both tumor stage and lymph node status were associated with survival. By contrast, others have noted that since approximately 90% of patients with SCLC present with either locally advanced or metastatic disease,⁷ the simpler VALSG system is sufficient to guide most treatment. A 2013 guideline from the American College of Chest Physicians recommended that SCLC be staged using both systems.⁹

Pretreatment Planning

At diagnosis, the majority of SCLC patients already have extensive disease. In general, patients staged as LD receive chemotherapy and radiation with remission as the goal. In contrast, SCLC patients with ED are not considered curable. In these cases, treatment primarily consists of palliative chemotherapy. Even without evidence of metastases in the brain, SCLC patients may also receive prophylactic cranial irradiation, as this has been demonstrated to prolong survival. Accurate staging of patients is necessary to select the treatment plan that will maximize a patient’s chances of survival. On the one hand, overstaging risks denying the patient potentially life-saving treatment; in contrast, understaging risks subjecting the patient to unnecessary risk of complications from more aggressive treatment. Timely diagnosis and staging is also important to maximize

survival: performing potentially unnecessary tests during the diagnostic and staging process could delay treatment initiation.

Imaging Tests

This section discusses several types of imaging tests used in the process of staging SCLC.

MDCT

Computed tomography scanners acquire cross-sectional images of the body using x-rays. While early CT scanners could only acquire one slice at a time, current state of the art is multi-detector CT (MDCT) scanners that can acquire as many as 640 slices at a time. These images can be created for viewing in any desired plane (multiplanar reconstruction). Since single-detector CT scanners are now obsolete, our report will exclude clinical studies using single-detector CT.

Intravenous injection of a radio-opaque contrast agent may be performed in some cases. Contrast-enhanced CT can provide additional information about the characteristics of a mass seen in the unenhanced scan which may facilitate characterization of the mass as probably malignant or probably benign.

MDCT has general strengths of widespread availability, high spatial resolution, and high speed, and is particularly useful for evaluation of the lungs, airways, bowel, and cortical bone. However, as it is a structural imaging modality, it may not detect early metastatic disease involving sites such as the bone marrow or lymph nodes, and is not always able to characterize lesions as benign or malignant based on their morphological properties. In addition, some patients cannot receive iodinated contrast material due to renal insufficiency, limiting evaluation for presence of hilar lymphadenopathy, vascular abnormalities, and lesion characterization. One potential concern about MDCT is patient exposure to ionizing radiation. Given the poor prognosis for SCLC when it is typically diagnosed, however, this concern is relatively unimportant.

MDCT of the chest is often one of the first tests performed to diagnose possible lung cancer. For staging SCLC, additional MDCT images are taken of abdomen, pelvis, or head to look for distant metastases.

PET/CT

Positron emission tomography (PET) is an imaging modality that localizes the uptake of a positron-emitting radioisotope within the body. [¹⁸F]-fluorodeoxyglucose (FDG) is the most commonly used PET radiotracer. Because FDG-PET identifies anatomic sites that harbor metabolically active malignant areas, FDG-PET is helpful to distinguish malignant tumors from benign nodules or masses. FDG-PET can also detect metabolically active metastases that may not be detected by structural imaging modalities (e.g., MDCT, MRI).

Because PET images lack anatomic detail, combined PET/CT scanners have been developed so the molecular information from PET can be anatomically localized with CT. As of 2014, PET without a concurrent CT is rarely performed. Therefore, our analysis of PET data will only include studies where patients were imaged using a combined PET/CT scanner.

FDG-PET/CT is sometimes available (but less often than CT), has high contrast resolution, and has high sensitivity for detection of sites of malignancy. Because it is quantitative, measurements of individual lesions or global disease burden may be performed. However, FDG uptake within a lesion does not always indicate malignancy. Infectious and non-infectious inflammatory lesions may also demonstrate FDG uptake, potentially leading to false positive results. Furthermore, some small or indolent malignant lesions (such as lung adenocarcinoma in situ or carcinoid tumor) may have little to no radiotracer uptake, potentially leading to false negative results. Lastly, patients with elevated serum glucose levels (greater than 200 mg/dl) do not generally undergo FDG-PET/CT, given the potential for false negative results. Again, one potential concern about MDCT is patient exposure to ionizing radiation. Given the poor prognosis for SCLC when it is typically diagnosed, however, this concern is relatively unimportant.

Discussions with key informants (KIs) suggested that there may be variation in practice among clinicians with regard to the use of FDG-PET/CT; some clinicians have replaced bone scintigraphy with FDG-PET/CT, while others continue to use bone scintigraphy in addition to PET/CT. Evidence on the comparative effectiveness of the modalities could help ascertain the clinical value of using both tests.

MRI

Magnetic resonance imaging (MRI) uses magnetic fields and radio waves to generate three-dimensional images of the body. Unlike PET and CT, MRI does not use ionizing radiation, and thus poses no radiation-related risks to the patient. Paramagnetic contrast agents can be administered during the MRI examination to give additional information about the nature of a mass in the same way as contrast agents can enhance a CT scan.

MRI is a structural and functional imaging technique with widespread availability, high spatial resolution, and high soft tissue contrast resolution; this imaging modality is particularly useful for detection and characterization of lesions within tissues even when subcentimeter in size, as well as for evaluation of the internal architecture of organs/tissues such as the brain, spinal cord, breasts, bone marrow, muscles, tendons, ligaments, cartilage, and other solid organs in the body. Also, functional imaging capabilities such as diffusion-weighted imaging and magnetic resonance spectroscopy may be used to improve diagnostic accuracy. MRI examinations take longer to perform and generally cost more than MDCT, and some patients with implanted electronic or metallic devices or with claustrophobia cannot undergo MRI.

In practice, most KIs agreed that during the staging process, all patients receive an MRI of the brain (unless contraindications exist); however, MRI of the chest is rarely part of the staging process.

PET/MRI

Combined PET and MRI scanners are a recent technical development: they promise the sensitivity of PET combined with the anatomic detail of MRI. PET/MRI is a hybrid molecular/structural imaging technique that possesses the general strengths provided by PET and MRI listed above which can be obtained in a single examination, along with a lower radiation dose compared to PET/CT, and potentially improved PET quantification and motion compensation. However, in addition to the weaknesses of PET and MRI listed above, it is not widely available, not currently reimbursed by insurance companies,

more expensive in terms of instrumentation than PET/CT, MDCT, and MRI, involves longer examination times, and requires additional training of personnel in terms of safety, protocol optimization, and study interpretation. KIs agreed that this technology is in the early state of commercialization and not yet widely used.

Endobronchial Ultrasound (EBUS) and Endoscopic Ultrasound (EUS)

EBUS is a bronchoscopic technique utilizing ultrasonography to visualize structures within and adjacent to the airway wall, whereas EUS is an endoscopic technique that utilizes ultrasonography to visualize structures within and adjacent to the esophageal wall. These techniques are minimally invasive and can be performed on an outpatient basis. Patients with suspected spread of lung cancer to mediastinal lymph nodes may undergo preoperative (or intraoperative) EBUS guided biopsy and/or EUS guided biopsy for the purpose of real-time minimally invasive pathologic mediastinal N staging, even for lymph nodes that are subcentimeter in size or located near blood vessels. EBUS is used to sample hilar and mediastinal lymph nodes that surround the tracheal bronchial tree, whereas EUS can be used to sample other lymph nodes in proximity to the esophagus. These techniques are utilized during the staging workup of patients with lung cancer (most often non-small cell lung carcinoma (NSCLC)), as accurate N staging is important to determine the optimal therapeutic approach (i.e., surgical vs. non-surgical therapy), especially since the sensitivity and specificity of CT, MRI, and FDG-PET/CT for non-invasive N staging of lung cancer are not 100%.

EBUS-guided transbronchial needle aspiration (EBUS-TBNA) is generally performed if suspected lymph nodes are in the anterior or superior mediastinum and appear to be accessible based on prior cross-sectional imaging, whereas EUS-guided fine needle aspiration (EUS-FNA) may initially be utilized for nodes that are paraesophageal or subaortic in location or located in the posterior or inferior mediastinum. EBUS-TBNA can also be used to sample hilar lymph nodes.

EBUS can also be used to inform T staging of lung cancer, and may potentially improve the determination of the distance between the tumor and the carina, which is a determinant of T stage. Furthermore, it may help to distinguish between airway invasion by tumor compared to airway compression by tumor. EBUS is also used for diagnostic evaluation of endobronchial lesions, peripherally located pulmonary nodules, mediastinal lesions, as well as for guiding endobronchial therapy.

Bone Scintigraphy

Bone scintigraphy uses a gamma camera to create two-dimensional images of the distribution of a radiotracer, typically technetium-99m methylene diphosphonate (Tc99m-MDP). Scintigraphy of radiotracers that localize to the bones is useful in cancer staging because areas of high radiotracer uptake could represent metastases; however, uptake occurs with other common conditions as well, meaning that false positive findings are not unusual.

Bone scintigraphy is a planar molecular imaging technique with widespread availability, high contrast resolution, and relatively low cost compared to FDG-PET/CT. However, false negative results can occur, particularly when lytic osseous lesions are present, as

bone scintigraphy indirectly reveals sites of metabolic bone turnover in reaction to bone marrow metastatic disease rather than the metastatic lesions directly. False positive results can also occur due to other non-malignant conditions that lead to increased bone turnover, such as fractures and osteomyelitis. Also, this technique only has planar capability (compared to tomographic capability of the other imaging techniques described above), further limiting its diagnostic capability. Lastly, bone scintigraphy is less quantitative compared to PET/CT, and does not allow evaluation of non-osseous structures of the body (in contrast to MDCT, MRI and PET).

As previously noted, KIs suggested there may be some variation in practice, as some clinicians no longer use bone scintigraphy as part the staging process.

Patient subgroups of particular interest

Currently, we are unaware of patient factors which might differentially affect the accuracy of imaging modalities specifically for SCLC. As previously noted, performance of various imaging modalities may be impacted by comorbidities such renal insufficiency which potentially limits the use of contrast for MDCT. Generally speaking, body habitus may limit the diagnostic quality and accuracy for any imaging modality. Many scanners are unable to safely accommodate patients above a particular weight. However, these are general patient considerations, not specific to the use of imaging for SCLC. Input from our KIs confirmed that aside from these general patient factors, there are no known patient factors which we should consider. We plan to consider studies which report on patient factors such as comorbidities and body habitus, if any such studies exist. We also plan to consider whether tumor characteristics may be associated with comparative accuracy and/or effectiveness.

Recent Guidance from Professional Societies and Need for Future Work

A 2013 guideline from the American College of Chest Physicians recommended that patients with either proven or suspected SCLC undergo CT of the chest and abdomen, or CT of the chest extending through the liver and adrenal glands, as well as MRI of the brain and bone scintigraphy.⁹ Furthermore, in patients thought to have limited disease, the guideline suggested FDG-PET/CT imaging to improve detection of metastases (with the potential to increase the stage and change the treatment choice). However, the evidence underlying this recommendation of PET for potential upstaging is limited and considered weak. The last search date of the guideline was 2011; therefore almost three years of additional literature will be available (2012 through October 2014 when the searches will be performed). A primary contribution of this report will be to update the evidence base by including more recent studies.

Similarly, in 2014, the American College of Radiology (ACR) appropriateness criteria review gave the highest rating of “usually appropriate” (with regard to staging SCLC) to the following specific modalities: CT of the chest and abdomen with contrast, MRI of the head with and without contrast, and FDG-PET/CT from skull base to mid-thigh.¹⁰ Bone scintigraphy was rated as “may be appropriate” and considered unnecessary if PET/CT had been performed. The ACR noted that PET/CT is often helpful in staging of SCLC, and may result in a change in staging in up to 17% of cases: mostly cases where PET/CT detects extensive disease that was not detected by other modalities. It may also detect

additional involved lymph nodes, leading to revisions in treatment plans for patients scheduled to receive radiotherapy. As PET/CT has already become frequently adopted as part of the SCLC staging process, it is important to establish whether or not evidence exists to support this practice.

II. The Key Questions

Question 1: What are the test concordance and comparative accuracy of imaging tests (MDCT, PET/CT, MRI, PET/MRI, EBUS, EUS, bone scintigraphy) for the pretreatment staging of small cell lung cancer?

- a. Test concordance
- b. Sensitivity
- c. Specificity
- d. Positive Predictive Value
- e. Negative Predictive Value
- f. Positive Likelihood Ratio
- g. Negative Likelihood Ratio

Question 2: When used for the pretreatment staging of small cell lung cancer, what is the comparative effectiveness of imaging tests (MDCT, PET/CT, MRI, PET/MRI, EBUS, EUS, bone scintigraphy) on later outcomes?

- a. Choice of treatment (e.g., surgery, chemotherapy, radiation)
- b. Timeliness of treatment
- c. Tumor response
- d. Harms due to overtreatment or undertreatment
- e. Survival
- f. Quality of life

Question 3. To what extent are the following factors associated with the comparative accuracy or effectiveness of imaging tests (MDCT, PET/CT, MRI, PET/MRI, EBUS, EUS, bone scintigraphy) when used for the pretreatment staging of small cell lung cancer?

- a. comorbidities
- b. body habitus
- c. tumor characteristics

Population(s)

- Adults with diagnosed SCLC or combined SCLC

Interventions

- Any of the following imaging tests when used for pretreatment staging:
 - MDCT

- PET/CT
- MRI
- PET/MRI
- EBUS
- EUS
- Bone scintigraphy

Comparators

- Single test (one of the above) vs. single test (another one of the above)
- Single test (one of the above) vs. single test (a specific variant of the same modality)
- Single test (one of the above) vs. multiple tests (more than one of the above)
- Multiple test (more than one of the above) vs. other multiple tests (more than one of the above)
- Test comparisons for patients with comorbid illnesses vs. those without (KQ3)
- Test comparisons at different levels of body habitus (KQ3)
- Test comparisons for different tumor characteristics (KQ3)

Outcomes

Intermediate outcomes

1. Test concordance (the percentage of patients for whom two imaging tests give the same result or different results)
2. Sensitivity (KQ1 and KQ3) (separately for different portions of the anatomy such as mediastinal lymph nodes, brain, etc.)
3. Specificity (KQ1 and KQ3) (separately for different portions of the anatomy such as mediastinal lymph nodes, brain, etc.)
4. Timeliness of treatment (KQ2 and KQ3)
5. Choice of treatment (KQ2 and KQ3)
6. Tumor response (KQ2 and KQ3)

Patient-centered outcomes

7. Survival (KQ2 and KQ3)
8. Quality of life (KQ2 and KQ3)
9. Harms due to overtreatment or undertreatment (KQ2 and KQ3)

Timing

- For test concordance: no minimum follow-up
- For accuracy: no minimum follow-up
- For timeliness of treatment, timing is the outcome itself
- For choice of treatment, no minimum follow-up
- For tumor response, no minimum follow-up

- For harms due to overtreatment or undertreatment, no minimum follow-up
- For survival and quality of life, at least six months minimum follow-up

Setting

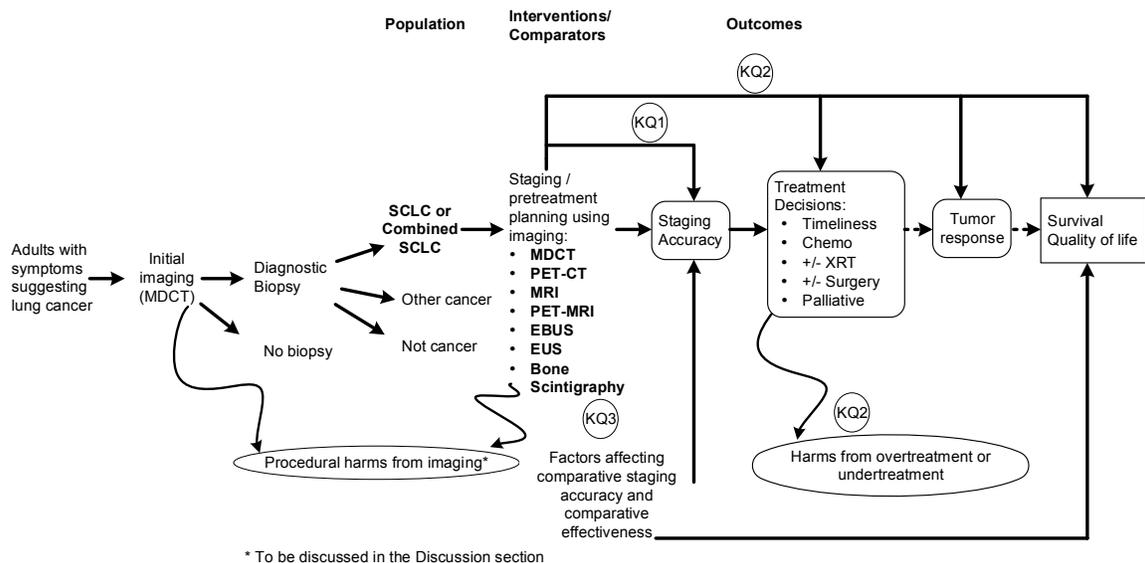
- Any setting

The Draft Key Questions were posted for public comment from 9/4/2014 to 9/24/2014. One commenter provided four comments, none of which engendered changes in the Key Questions or Analytic Framework. The commenter expressed concern that comparative effectiveness evaluations would not recognize the complimentary contribution of different imaging modalities to patient-level decision-making: our review is intended to inform patient-level decision-making, not replace it. During the posting period, the review team decided to expand the scope in the following ways:

- Add two additional technologies: EBUS and EUS. These invasive technologies are sometimes used in the staging of SCLC and therefore, in the interest of comprehensiveness, we plan to include them as staging modalities to evaluate.
- Add the concept of comparing variants of a given modality (e.g., diffusion-weighted MRI vs. non-diffusion-weighted MRI).
- Add tumor response as an effectiveness outcome. Since we are already planning to summarize any evidence on comparative accuracy and comparative long-term outcomes, it makes sense to look for the intermediate outcome of comparative tumor response.
- Add the concept of tumor characteristics to Key Question 3, since different characteristics may be associated with differential accuracy or effectiveness.

We edited the Key Questions and Analytic Framework accordingly.

III. Analytic Framework



IV. Methods

Criteria for Inclusion/Exclusion of Studies in the Review

Our inclusion criteria are listed below in five categories: Publication criteria, study design criteria, patient criteria, test criteria, and data criteria.

Publication criteria

- a. Full-length articles: The article must have been published as a full-length, peer-reviewed study.
- b. Redundancy: To avoid double-counting patients, in instances in which several reports of the same or overlapping groups of patients were available, only outcome data based on the larger number of patients will be included. However, we will include data from publications with lower numbers of patients when either (a) a publication with lower patient enrollment reported an included outcome that was not reported by other publications of that study, or (b) a publication with lower patient enrollment reported longer follow up data for an outcome.
- c. English language: Although we recognize that in some situations exclusion of non-English studies could lead to bias, we believe that the few instances in which this may occur do not justify the time and cost typically necessary for translating studies.
- d. Publication date: We will include studies published since January 1, 2000. Technical progress in all of the imaging modalities under consideration means that older studies are likely to underestimate the diagnostic performance of these modalities.

Study Design Criteria

- a. The study must have provided data comparing two or more tests (or test strategies). If direct comparisons are insufficient to determine comparative accuracy, we will broaden the inclusion criteria to include single-modality studies with an appropriate reference standard for determining sensitivity and specificity (e.g., a set of studies reporting the accuracy of MRI at mediastinal node staging of SCLC, compared to a separate set of studies reporting the accuracy of CT at mediastinal node staging of SCLC). This is an indirect comparison of modalities, so conclusions based on it will necessarily be weaker.
- b. For comparisons of variants of a given modality, and studies of patient factors or tumor characteristics for KQ3, the comparison must have been planned in advance.
- c. For comparative accuracy (KQ1), the study must have compared both tests to a common reference standard. The reference standard must not have been defined by either imaging test being assessed. We will set no requirements on what the reference standard must be (e.g., biopsy, clinical follow-up).
- d. For comparative effectiveness (KQ2), some patients must have received one of the imaging tests (or test strategies), and a separate group of patients must have received a different imaging test (or test strategy). This design permits a

comparison of how the choice of test (or test strategy) may influence choice of treatment, timeliness of treatment, harms due to over/under treatment, survival and/or quality of life.

- e. For the influence of patient factors or tumor characteristics (KQ3), the study must have reported comparative accuracy data stratified by patient/tumor factor, or comparative effectiveness data stratified by patient/tumor factor.

Patient criteria

- a. The study reported data specifically on patients undergoing staging for SCLC, or if the data were combined with other types of patients, at least 85% of the patients in the reported data were undergoing staging for SCLC.
- b. Adults. At least 85 percent of patients must have been aged 18 years or older, or data must have been reported separately for those aged 18 years or older.
- c. Studies of the staging of *primary* small cell lung cancer will be included. Testing for *recurrent* small cell lung cancer will be excluded.
- d. Data on imaging tests performed after any form of treatment (e.g., chemotherapy) were excluded, but pretreatment imaging data were considered.

Test criteria

- a. Type of test. Only studies of the imaging tests of interest were included (listed in the key questions above). Studies of computed tomography (CT) that did not explicitly state that (or it could not be determined that) CT was MDCT were assumed to be MDCT. Given our publication date criterion of 2000 and later, we believe it is safe to assume that CT performed in such studies was MDCT.
- b. PET/CT must have been based on a dedicated PET/CT machine. We excluded studies in which PET and CT were acquired separately and later fused for the following reasons. First, methods for image fusion can vary widely. Since accurate staging relies on precise localization of the margin of tumor or metastasis, inclusion of less precise means of combining PET and CT images would underestimate the performance of PET/CT. Second, image fusion is no longer the state of the art in PET/CT, and since this report is intended to guide use of imaging technologies in the future, it should reflect the technology in use at present.

Data criteria

- a. The study must have reported data pertaining to one of the outcomes of interest (see the Key Questions).
 - For test concordance (part of KQ1), this means reporting the number of patients for whom the two tests provided the same or different results.
 - For comparative accuracy (part of KQ1), this means reporting enough information for one to calculate both sensitivity and specificity, along with their corresponding confidence intervals (CIs).

- For comparative choice of treatment (part of KQ2), this means reporting the percentage of patients who received a specific treatment choice, for one test or test strategy compared to another test or test strategy.
 - For comparative timeliness of treatment (part of KQ2), this means reporting the duration of time elapsed before the initiation of treatment, for one test or test strategy compared to another test or test strategy.
 - For comparative tumor response (part of KQ2), this means reporting the percentage of patients whose tumor responded to treatment, for one test or test strategy compared to another test or test strategy.
 - For comparative harms of overtreatment or undertreatment (part of KQ2), this means reporting the percentage of patients who were overtreated or undertreated (based on authors' judgment), for one test/test strategy compared to another test or test strategy.
 - For comparative survival (part of KQ2), this means either reporting median survival after each imaging test or test strategy, or mortality rates at a given time point, or other patient survival such as a hazard ratio.
 - For quality of life (part of KQ2), this means reporting data on a previously tested quality-of-life instrument (such as the SF-36) separately for each imaging test or test strategy.
 - For patient factors or tumor characteristics (KQ3), this means reporting data on whether such factors are associated with either comparative accuracy or comparative effectiveness (or both).
- b. Regarding the minimum patient enrollment, we required data on at least 10 patients per imaging test or test strategy.
 - c. For all KQs, the reported data must have included at least 50 percent of the patients who had initially enrolled in the study.

Searching for the Evidence: Literature Search Strategies for Identification of Relevant Studies to Answer the Key Questions

Literature searches will be performed by medical librarians within the Evidence-Based Practice Center (EPC) Information Center; searches will follow established systematic review methods. We will search the following databases using controlled vocabulary and text words: EMBASE, MEDLINE, PubMed, and the Cochrane Library.

The following grey literature sources will be searched using text words: ClinicalTrials.gov, Centers for Medicare & Medicaid Medicare Coverage Database, ECRI Institute Health Devices, Healthcare Standards, Internet, Medscape, National Guideline Clearinghouse™, and the U.S. Food and Drug Administration.

An example search strategy is shown in Appendix A.

Literature screening will be performed by experienced research analysts using the database Distiller SR (Evidence Partners, Ottawa, Canada). Literature search results will initially be screened for relevancy. Relevant abstracts will be dual screened against the

study inclusion criteria in duplicate. Studies that appear to meet the study inclusion criteria will be retrieved in full and screened in duplicate against the study inclusion criteria. All disagreements will be resolved by consensus discussion among the two original screeners and an additional third screener.

The literature searches will be updated during the peer-review process before finalization of the review.

Data Abstraction and Data Management

Abstraction forms will be constructed in Microsoft Excel (Microsoft Corporation, Redmond, Washington), and the data will be abstracted into these forms. Elements to be abstracted include general study characteristics (e.g., country, setting, study design, enrolled N, funding source), patient characteristics (e.g., age, sex, body habitus, comorbidities), tumor characteristics, details of the imaging methodology (e.g., radiotracer, timing of test), risk of bias items, and outcome data. Comparative outcome data will be dual abstracted to ensure accuracy, with discrepancies resolved by consensus. Multiple publications of the same study will be grouped together as a single study. Such multiple publications will be identified by examination of author names, study centers, patient enrollment dates, and imaging technologies.

Assessment of Methodological Risk of Bias of Individual Studies

For studies of comparative accuracy (i.e., those for Key Question 1), we will assess the risk of bias using ten items (see Appendix B). Some of these items will be from the QUADAS-2 instrument (Whiting et al. 2011),¹¹ and others will be created for this report. An example item that we may use in this report is “Was prior experience with the test (technicians, readers) similar for the two imaging tests being compared in the study?” This is important because if readers are simply more experienced with one test than another, an observed difference in accuracy may be due to reader experience rather than the tests themselves.

For studies addressing comparative effectiveness (i.e., those for Key Question 2), we will use an internal validity rating scale for comparative studies to assess the risk of bias of each individual study. This instrument was developed by ECRI Institute¹² with respect to the impact of study design on bias in comparative studies and is consistent with the guidance in the Methods Guide for Comparative Effectiveness Reviews (see Appendix B).¹³ Each question in the instrument addresses an aspect of study design or conduct that can help protect against bias, such as randomization of group assignment, or blinding outcome assessors to patient group assignment. Each question can be answered “yes,” “no,” or “not reported,” and each is phrased such that an answer of “yes” indicates that the study reported a protection against bias on that aspect.

For studies addressing the association between various factors (comorbidities, body habitus, tumor characteristics) and either comparative accuracy or comparative effectiveness (i.e., those for Key Question 3), we will use one of the above two instruments (either for comparative accuracy or comparative effectiveness), as appropriate for the outcome.

For each outcome, studies will be rated as “Low,” “Medium,” or “High” risk of bias. A “Low” rating will be defined by selecting critical questions from the rating scale that

must be answered as “yes.” For example, for a comparative study to be rated as “Low” risk of bias, it may need to have randomly assigned participants, adequately concealed allocation, demonstrated good baseline comparability between groups on the outcomes, and blinded outcome assessors. Conversely, a “High” rating will be defined by selecting critical questions that may be answered “no” or “not reported”.

Studies reporting multiple outcomes may be assigned one risk-of-bias category for some outcomes but another risk-of-bias category for other outcomes because the risk of bias can depend on the outcome being measured (e.g., higher rates of missing data for some outcomes).

Data Synthesis

Decisions about whether meta-analysis is appropriate will depend on the judged clinical homogeneity of the different study populations, imaging and treatment protocols, and outcomes. When meta-analysis is not possible (due to limitations of reported data) or is judged to be inappropriate, the data will be synthesized using a descriptive narrative review approach.

For Key Question 1 on comparative accuracy, we will synthesize sensitivity and specificity of each test on its own using a bivariate mixed-effects binomial regression model as described by Harbord et al.¹⁴ All such analyses will be computed by the STATA 13 statistical software package using the “midas” command.¹⁵ In cases in which a bivariate binomial regression model cannot be fit, we will meta-analyze the data using a random-effects model and the software package Meta-Disc.¹⁶ To compare two tests, we will then use the EPC methods described in equation 39 of Trikalinos et al. (2013).¹⁷ If an indirect analysis is attempted, we will investigate appropriate methods for quantitative synthesis.

For Key Question 2 on comparative effectiveness, if meta-analysis is appropriate, we will compute effect sizes and measures of variance using standard methods, and will perform Knapp-Hartung random-effects meta-analysis. If heterogeneity is encountered, we will use meta-regression to explore possible causes. Again, if we attempt an indirect analysis, we will investigate appropriate methods.

For Key Question 3 on patient factors and tumor characteristics, we will calculate for each study the association between a patient factor (e.g., age) and comparative accuracy or effectiveness (e.g., the difference in sensitivity between two tests). If appropriate, we will meta-analyze the correlations using standard techniques and Knapp-Hartung standard errors.

Grading the Strength of Evidence (SOE) for Major Comparisons and Outcomes

We will determine the strength of evidence grade for the following outcomes:

- Test concordance (Key Question 1)
- Comparative accuracy (Key Question 1) (this is comprised of both sensitivity and specificity)
- Comparative timeliness of treatment (Key Question 2)
- Comparative choice of treatment (Key Question 2)
- Comparative tumor response (Key Question 2)
- Comparative survival (Key Question 2)
- Comparative quality of life (Key Question 2)
- Comparative harms due to overtreatment (Key Question 2)
- Comparative harms due to undertreatment (Key Question 2)
- Association between age and comparative accuracy (Key Question 3)
- Association between body habitus and comparative accuracy (Key Question 3)
- Association between tumor characteristics and comparative accuracy (Key Question 3)
- Association between age and comparative effectiveness (Key Question 3)
- Association between body habitus and comparative effectiveness (Key Question 3)
- Association between tumor characteristics and comparative effectiveness (Key Question 3)

Each grade will be determined separately for each modality comparison (e.g., CT vs PET/CT). The reason that we will consider sensitivity and specificity together for a single grade is that the two measures are mathematically related (i.e., a less stringent threshold for declaring a case a positive will necessarily improve sensitivity and will necessarily worsen specificity). The reason that we will not separately grade other accuracy-related outcomes (i.e., predictive values or likelihood ratios) is that these grades would be redundant with the grades for accuracy, because our estimates would be based on the same studies and the same quantitative syntheses.

We will use the EPC system for grading evidence on diagnostic tests as described in the EPC guidance chapter by Singh et al. (2012).¹⁸ This system uses up to eight domains as inputs (risk of bias, directness, consistency, precision, publication bias, dose-response association, all plausible confounders would reduce the effect, and strength of association). The output is a grade for the strength of evidence: High, Moderate, Low, or Insufficient. This grade is provided separately for each outcome of each comparison of each Key Question.

A grade of Insufficient will be given when the evidence does not permit a conclusion for the outcome of interest and the two modalities being compared. For example, if the outcome is test accuracy, and the comparison is CT to PET/CT, the evidence may permit a conclusion that either 1) CT is more accurate, 2) PET/CT is more accurate, or 3) the tests are similarly accurate. If none of these three conclusions can be drawn, then evidence is graded Insufficient for that comparison. In order to draw a conclusion that two modalities differ for an outcome, we will use a p value less than 0.05 two-tailed (i.e., the standard value for alpha). In order to draw a conclusion that two modalities are approximately equivalent for an outcome, we will use the independent judgment of two analysts (with disagreements resolved by discussion).

If the evidence is sufficient to permit a conclusion, then the grade is deemed High, Moderate, or Low. The grade will be provided by two independent raters, and discrepancies will be resolved by consensus. Below, we discuss the eight domains and how they will be considered as inputs to the grade:

Risk of bias (see the section Assessment of Methodological Risk of Bias of Individual Studies above). If the evidence permits a conclusion, and all else being equal, a set of studies at Low risk of bias yield a higher strength of evidence grade than a set of studies at Moderate or High risk of bias.

Directness. Our initial inclusion criteria require direct comparisons between modalities. If we attempt indirect comparisons, then the evidence will be downgraded accordingly. This is because different sets of studies may have enrolled slightly different types of patients or used slightly different reference standards for determining patients' true SCLC stages.

Consistency. Consistency among study results will be judged based on whether the studies' findings suggest the same direction of effect (e.g., that one test is better than another).

Precision. Precision of the combined study results will be judged by an assessment of the meta-analytic confidence interval, or if meta-analysis was not performed, by an assessment of the precision of the individual studies (based on confidence intervals or numbers of patients).

Reporting bias. This will be addressed by noting the presence of abstracts or ClinicalTrials.gov entries describing studies that did not subsequently appear as full published articles. If many such studies exist, this will tend to decrease the strength of the evidence. We will also consider the funding sources of the studies in determining the risk of reporting bias.

Dose-response association. This factor is used for studies of treatments, and is not relevant for this topic.

All plausible confounders would reduce the effect. This domain means that a set of studies may be biased *against* finding a difference between two modalities, and yet the studies still found an important difference. Thus, if the studies had controlled for the confounders, the effect would have been even larger. This domain will generally increase the strength of evidence grade.

Strength of association. While the Cochrane Collaboration and other reviewers have objective criteria for defining this domain in the context of treatments, there are not such criteria for diagnostic test effectiveness. Therefore, this domain will be judged by EPC team members based on whether the strength of the effect (e.g., the extent of difference in accuracy between two tests) is so large that the potential study biases could not explain it. If true, this domain will generally increase the strength of evidence grade.

Assessing Applicability

The applicability of the evidence involves four key aspects: patients, tests/interventions, comparisons, and settings. In considering the applicability of the findings to patients, we will consult large studies to ascertain the typical characteristics of patients newly given a diagnosis of pancreatic adenocarcinoma (e.g., age, gender) and then assess whether the included studies enrolled similar patients. Some aspects of interventions may also affect applicability, for example if a study uses an uncommonly used radiotracer. Settings of care will be described, and if data permit, subgroups of studies by setting will be analyzed separately.

V. References

1. American Cancer Society (ACS). Cancer facts & figures 2014. Atlanta (GA): American Cancer Society (ACS); 2014. 70 p. Also available: <http://www.cancer.org/acs/groups/content/%40research/documents/webcontent/acspc-042151.pdf>.
2. National Cancer Institute (NCI). Small cell lung cancer treatment (PDQ) health professional version. [internet]. Bethesda (MD): National Cancer Institute (NCI), National Institutes of Health (NIH); 2014 Feb 21 [accessed 2014 Jun 11]. [33 p]. Available: <http://www.cancer.gov/cancertopics/pdq/treatment/small-cell-lung/healthprofessional/page1>.
3. SEER Cancer Statistics Review, 1975-2011. Bethesda (MD): National Cancer Institute; 31 p.
4. American Cancer Society. Cancer prevention & early detection facts & figures 2013. Atlanta (GA): American Cancer Society; 2013. 65 p.
5. Micke P, Faldum A, Metz T, et al. Staging small cell lung cancer: Veterans Administration Lung Study Group versus International Association for the Study of Lung Cancer--what limits limited disease?. *Lung Cancer*. 2002 Sep;37(3):271-6. PMID: 12234695
6. Trends in lung cancer morbidity and mortality. Chicago (IL): American Lung Association; 2012 Apr. 34 p.
7. Glisson BS, Byers LA. Pathobiology and staging of small cell carcinoma of the lung. In: Lilenbaum RC, Nicholson A, Jett JR, editor(s). *UptoDate*. [internet]. Waltham (MA): UptoDate; 2013 May 07 [accessed 2014 Jun 12]. Available: <http://www.uptodate.com/home/index.html>.
8. Shepherd FA, Crowley J, Van Houtte P, et al. The International Association for the Study of Lung Cancer lung cancer staging project: proposals regarding the clinical staging of small cell lung cancer in the forthcoming (seventh) edition of the tumor, node, metastasis classification for lung cancer. *J Thorac Oncol*. 2007 Dec;2(12):1067-77. Also available: <http://dx.doi.org/10.1097/JTO.0b013e31815bdc0d>. PMID: 18090577
9. Jett JR, Schild SE, Kesler KA, et al. Treatment of small cell lung cancer: diagnosis and management of lung cancer, 3rd ed: American college of chest physicians evidence-based clinical practice guidelines. *Chest*. 2013 May;143(5):e400S-19S. Also available: <http://dx.doi.org/10.1378/chest.12-2363>. PMID: 23649448
10. Ravenel JG, Rosenzweig KE, Kirsch J, et al. ACR Appropriateness Criteria Non-invasive Clinical Staging of Bronchogenic Carcinoma. *J Am Coll Radiol*. 2014 Sep;11(9):849-56. Also available: <http://dx.doi.org/10.1016/j.jacr.2014.05.020>. PMID: 25086958

11. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011 Oct 18;155(8):529-36. PMID: 22007046
12. Treadwell JR, Tregear SJ, Reston JT, et al. A system for rating the stability and strength of medical evidence. *BMC Med Res Methodol*. 2006;6:52. Also available: <http://www.biomedcentral.com/1471-2288/6/52>. PMID: 17052350
13. Viswanathan M, Ansari MT, Berkman ND, et al. Assessing the risk of bias of individual studies in systematic reviews of health care interventions. [AHRQ Publication No. 12-EGC047-EF]. Rockville (MD): Agency for Healthcare Research and Quality (AHRQ); 2012 Mar. 33 p. (Methods guide for comparative effectiveness reviews; Also available: <http://www.effectivehealthcare.ahrq.gov>.)
14. Harbord RM, Deeks JJ, Egger M, et al. A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics*. 2007 Apr;8(2):239-51. PMID: 16698768
15. STATA statistics/data analysis. MP parallel edition. Ver. 10.0. College Station (TX): StataCorp; 1984-2007. Single user Stata for Windows. Also available: <http://www.stata.com>.
16. Zamora J, Abairra V, Muriel A, et al. Meta-DiSc: a software for meta-analysis of test accuracy data. *BMC Med Res Methodol*. 2006;6:31. PMID: 16836745
17. Trikalinos TA, Hoaglin DC, Small KM, et al. Evaluating practices and developing tools for comparative effectiveness reviews of diagnostic test accuracy: methods for the joint meta-analysis of multiple tests. Methods research report. (Prepared by the Tufts Evidence-based Practice Center, under Contract No. 290-2007-10055-I.) AHRQ publication no. 12(13)-EHC151-EF. Rockville (MD): Agency for Healthcare Research and Quality (AHRQ); 2013 Jan. 49 p. Also available: http://effectivehealthcare.ahrq.gov/ehc/products/291/1380/Methods%20Report_Evaluating-Practices-Developing-Tools_Final_01-14-2013.pdf.
18. Singh S, Chang S, Matchar DB, et al. Grading a body of evidence on diagnostic tests (AHRQ publication no. 12-EHC079-EF). Chapter 7 of Methods guide for medical test reviews (AHRQ publication no. 12-EHC017). Rockville (MD): Agency for Healthcare Research and Quality (AHRQ); 2012 Jun 1. Also available: <http://www.effectivehealthcare.ahrq.gov/reports/final.cfm>.

VI. Definition of Terms

Sensitivity: the percentage of patients with advanced disease who are correctly classified by the test as having advanced disease

Specificity: the percentage of patients with limited disease who are correctly classified by the test as having limited disease

Positive Predictive Value: the percentage of patients who are classified by the test as having advanced disease who actually have advanced disease

Negative Predictive Value: the percentage of patients who are classified by the test as having limited disease who actually have limited disease

Positive Likelihood Ratio: Sensitivity divided by 1-specificity. This ratio can be applied to the pretest probability of disease to calculate the posttest probability of disease.

Negative Likelihood Ratio: 1-Sensitivity divided by specificity. This ratio can be applied to the pretest probability of disease to calculate the posttest probability of disease.

VII. Summary of Protocol Amendments

If we need to amend this protocol, we will give the date of each amendment, describe the change and give the rationale in this section. Changes will not be incorporated into the protocol. Example table below:

Date	Section	Original Protocol	Revised Protocol	Rationale

VIII. Review of Key Questions

AHRQ posted the key questions on the Effective Health Care Website for public comment from September 4th, 2014 to September 24th, 2014. The EPC refined and finalized the key questions after review of the public comments, and input from Key Informants and the Technical Expert Panel (TEP). This input is intended to ensure that the key questions are specific and relevant.

IX. Key Informants

Key Informants are the end users of research, including patients and caregivers, practicing clinicians, relevant professional and consumer organizations, purchasers of health care, and others with experience in making health care decisions. Within the EPC program, the Key Informant role is to provide input into identifying the Key Questions for research that will inform healthcare decisions. The EPC solicits input from Key Informants when developing questions for systematic review or when identifying high priority research gaps and needed new research. Key Informants are not involved in analyzing the evidence or writing the report and have not reviewed the report, except as given the opportunity to do so through the peer or public review mechanism.

Key Informants must disclose any financial conflicts of interest greater than \$10,000 and any other relevant business or professional conflicts of interest. Because of their role as end-users, individuals are invited to serve as Key Informants and those who present with potential conflicts may be retained. The TOO and the EPC work to balance, manage, or mitigate any potential conflicts of interest identified.

X. Technical Experts

Technical Experts constitute a multi-disciplinary group of clinical, content, and methodological experts who provide input in defining populations, interventions, comparisons, or outcomes and identify particular studies or databases to search. They are selected to provide broad expertise and perspectives specific to the topic under development. Divergent and conflicting opinions are common and perceived as health scientific discourse that results in a thoughtful, relevant systematic review. Therefore study questions, design, and methodological approaches do not necessarily represent the views of individual technical and content experts. Technical Experts provide information to the EPC to identify literature search strategies and recommend approaches to specific issues as requested by the EPC. Technical Experts do not do analysis of any kind nor do they contribute to the writing of the report. They have not reviewed the report, except as given the opportunity to do so through the peer or public review mechanism.

Technical Experts must disclose any financial conflicts of interest greater than \$10,000 and any other relevant business or professional conflicts of interest. Because of their unique clinical or content expertise, individuals are invited to serve as Technical Experts and those who present with potential conflicts may be retained. The TOO and the EPC work to balance, manage, or mitigate any potential conflicts of interest identified.

XI. Peer Reviewers

Peer reviewers are invited to provide written comments on the draft report based on their clinical, content, or methodological expertise. The EPC considers all peer review comments on the draft report in preparation of the final report. Peer reviewers do not participate in writing or editing of the final report or other products. The final report does not necessarily represent the views of individual reviewers. The EPC will complete a disposition of all peer review comments. The disposition of comments for systematic reviews and technical briefs will be published three months after the publication of the evidence report.

Potential Peer Reviewers must disclose any financial conflicts of interest greater than \$10,000 and any other relevant business or professional conflicts of interest. Invited Peer Reviewers may not have any financial conflict of interest greater than \$10,000. Peer reviewers who disclose potential business or professional conflicts of interest may submit comments on draft reports through the public comment mechanism.

XII. EPC Team Disclosures

EPC core team members must disclose any financial conflicts of interest greater than \$1,000 and any other relevant business or professional conflicts of interest. Related financial conflicts of interest that cumulatively total greater than \$1,000 will usually disqualify EPC core team investigators.

XIII. Role of the Funder

This project was funded under Contract No. xxx-xxx from the Agency for Healthcare Research and Quality, U.S. Department of Health and Human Services. The Task Order Officer reviewed contract deliverables for adherence to contract requirements and quality. The authors of this report are responsible for its content. Statements in the report should not be construed as endorsement by the Agency for Healthcare Research and Quality or the U.S. Department of Health and Human Services.

Appendix A: Example Search Strategy

English language, 2000-2014

EMBASE syntax

EMBASE/Medline

Set Number	Concept	Search statement
1	Small cell lung cancer	small cell lung carcinoma'/exp OR 'carcinoma small cell'/de OR 'lung small cell cancer':de
2		('small-cell' OR 'small cell') NEAR/2 (lung OR bronch*) OR 'oat cell' OR sclc
3		2 NOT ('non-small cell':ti OR 'non-small-cell':ti OR 'non small cell':ti OR 'nonsmall cell':ti OR nsccl:ti)
4		#1 OR # 3
5	Lung Symptoms	'respiratory tract disease'/exp OR 'respiratory tract disease' OR 'lung disease'/exp OR 'lung tumor'/exp
6		(lung* OR pulmonary OR bronch* OR chest):ti AND (age* OR smok* OR symptom* OR wheez* OR cough* OR edema OR fibrosis OR asthma OR 'shortness of breath' OR pain* OR nodule*):ab,ti
7		#5 OR #6
8	Imaging	'computer assisted tomography'/exp OR 'emission tomography'/de OR 'nuclear magnetic resonance imaging'/exp OR 'diagnostic imaging'/exp OR 'tomography'/de OR 'respiratory-gated imaging' OR 'bone scintiscanning'
9		sdet OR mdct OR mri OR ct OR cat OR pet OR fdg NEAR/1 pet OR 'computed tomography' OR 'positron emission' OR 'magnetic resonance' OR 'multislice' OR 'multi slice' OR bone NEXT/2 scan*
10		'endosonography' /exp OR 'endosonography' OR 'eus' OR endoscop* NEXT/1 (ultrasound OR ultrasonography* OR echograph*)
11		'endobronchial echography'/exp OR 'endobronchial echography' OR 'endobronchial ultrasound'/exp OR 'endobronchial ultrasound' OR 'ebus'/exp OR 'ebus' OR endobronch* NEXT/1 (ultrasound OR ultrasonography* OR echograph*)
12		#8 OR #9 OR #10 OR #11
13	Pretreatment staging and	'cancer classification'/exp OR 'cancer classification' OR 'cancer staging'/exp OR 'cancer staging'

Set Number	Concept	Search statement
14	planning	'cancer'/de OR cancer OR tumor* OR tumour* OR mass* OR neoplasm* AND (stage OR 'staging'/de OR staging OR class*)
15		cancer:ab,ti OR neoplasm:ab,ti AND (care:ab,ti OR treatment:ab,ti OR therapy:ab,ti OR pretreatment:ab,ti OR 'pre-treatment':ab,ti OR plan:ab,ti)
16		#13 OR #14 OR #15
17	Prognosis	cancer prognosis'/exp OR 'cancer prognosis' OR prognos* OR predict* OR outcome* OR 'survival'/de OR survival OR 'quality of life'/de OR 'quality of life' OR qol
18	Combine Small Cell Lung Cancer AND Imaging	#4 AND #12
19	Combine Lung Symptoms AND Imaging	#7 AND #12
20	Combine SCLC AND Imaging AND Pretreatment staging and planning (KQ 1)	#16 AND #18
21	Combine Lung Symptoms AND Imaging AND Pretreatment staging and planning (KQ 1)	#16 AND #19
22	Combine SCLC AND Imaging AND Pretreatment staging and planning AND prognosis (KQ 2)	#17 AND #20

Set Number	Concept	Search statement
23	Combine Lung Symptoms AND Imaging AND Pretreatment staging and planning AND prognosis (KQ 2)	#17 AND #21
24	Combine final sets	#20 OR #21 OR #22 OR #23
25	Apply limits	#24 AND [humans]/lim AND [2000-2014]/py
26	Limit by publication and study type	#25 AND ('clinical article'/exp OR 'clinical article' OR 'clinical trial'/exp OR 'clinical trial' OR 'controlled study'/exp OR 'controlled study' OR 'intermethod comparison'/exp OR 'intermethod comparison' OR 'major clinical study'/exp OR 'major clinical study' OR 'retrospective study'/exp OR 'retrospective study')
27	guidelines	#25 AND ('practice guideline'/exp OR 'practice guideline' OR 'professional standard':de OR 'practice parameter' OR 'position statement' OR 'position paper' OR 'policy statement' OR standard*:ti OR guideline*:ti OR 'white paper' OR 'clinical pathway'/exp OR 'clinical pathway' OR 'clinical guideline' OR 'consensus development'/exp OR 'consensus development')
28	Systematic reviews	#25 AND ('research synthesis' OR pooled OR 'systematic review'/de OR 'meta analysis'/de OR ('evidence base' OR 'evidence based' OR methodol* OR systematic OR quantitative* OR studies OR search* AND ('review'/de OR review/it)))
29	Remove unwanted publication types	#25 NOT ('conference abstract'/it OR 'conference paper'/it OR 'editorial'/it OR 'letter'/it OR 'note'/it)
30	Combine sets	#26 OR #27 OR #28 OR #29

EMBASE.com Syntax:

* = truncation character (wildcard)

NEAR/*n* = search terms within a specified number (*n*) of words from each other in any order

- NEXT/*n* = search terms within a specified number (*n*) of words from each other in the order specified
- / = search as a subject heading
- exp = “explodes” controlled vocabulary term (e.g., expands search to all more specific related terms in the vocabulary’s hierarchy)
- mj = denotes a term that has been searched as a major subject heading
- :de = search in the descriptors field
- :lnk = floating subheading
- :it,pt. = source item or publication type
- :ti. = limit to title
- :ti,ab. = limit to title and abstract fields

PUBMED (PreMEDLINE)

Set Number	Concept	Search statement
1	Small cell lung cancer	(('small-cell' OR 'small cell') AND (lung OR bronch*)) OR 'oat cell' OR sclc
2		2 NOT ('non-small cell'[ti] OR 'non-small-cell'[ti] OR 'non small cell'[ti] OR 'nonsmall cell'[ti] OR nsclc[ti])
3	Lung Symptoms	(lung*[ti] OR pulmonary[ti] OR bronch*[ti] OR chest[ti]) AND (age* OR smok* OR symptom* OR wheez* OR cough* OR edema OR fibrosis OR asthma OR 'shortness of breath' OR pain* OR nodule*)
4	Imaging	sdct OR mdct OR mri OR ct OR (cat AND scan*) OR pet OR (fdg AND pet) OR 'computed tomography' OR 'positron emission' OR 'magnetic resonance' OR 'multislice' OR 'multi slice' OR “bone scan” OR “bone scintigraphy” OR (bone AND scan*)
5		'endosonography' OR 'eus' OR (endoscop* AND (ultrasound OR ultrasonography* OR echograph*))
6		'endobronchial echography' OR 'endobronchial ultrasound' OR 'ebus' OR (endobronch* AND (ultrasound OR ultrasonography* OR echograph*))
7		#4 OR #5 OR #6
8	Pretreatment staging and	(cancer OR tumor* OR tumour* OR mass* OR neoplasm*) AND (stage OR staging OR class*)

Set Number	Concept	Search statement
9	planning	(cancer[tiab] OR neoplasm[tiab] OR tumor*[tiab] OR tumour*[tiab]) AND (care[tiab] OR treatment[tiab] OR therapy[tiab] OR pretreatment[tiab] OR 'pre-treatment'[tiab] OR plan[tiab])
10		#8 OR #9
11	Prognosis	'cancer prognosis' OR prognos* OR predict* OR outcome* OR survival OR 'quality of life' OR qol
12	Combine SCLC OR Lung Symptoms	#2 OR #3
13	Combine SCLC OR Lung Symptoms with imaging	#7 AND #12
14	Combine SCLC OR Lung Symptoms with imaging and pretreatment planning	#10 AND #13
15	Combine with prognosis	#11 AND #14
16	In process	#15 AND (pubmednotmedline[sb] OR inprocess[sb] OR [publisher[sb]])
17	English	#16 AND English[la]

PubMed syntax:

- * = truncation character (wildcard)
- [ti] = limit to title field
- [tiab] = limit to title and abstract fields
- [tw] = text word

Appendix B. Instruments for Assessment of Methodologic Risk of Bias of Individual Studies

Instrument for assessing the risk of bias of comparative accuracy studies

1. Did the study enroll all, consecutive, or a random sample of patients?
2. Was the study unaffected by spectrum bias (e.g., patients with known status before the study, or patients selected for being difficult to diagnose/stage)?
3. Was prior experience with the test (technicians, readers) similar for the two imaging tests being compared in the study?
4. Were the imaging tests performed within one month of each other (to avoid the possibility that the patient's true condition changed between tests)?
5. Was knowledge of the other test complementary (either both tests were read with knowledge of the other results, or neither test was read with knowledge of the other)?
6. Did the interpreters have the same other information available at the time of interpretation for the two imaging tests (other clinical information, 3rd test results)?
7. Was each test's accuracy measuring using the same reference standard (or a similar proportion of patients who underwent different reference standards such as clinical follow-up and surgical findings)?
8. Were readers of both tests of interest blinded to the results of the reference standard (or the reference standard was unknowable until after the tests were read)?
9. Were the people determining the reference standard unaware of the diagnostic test results?

ECRI Instrument for comparative studies

1. Were patients randomly assigned to the study's groups?
2. Did the study use appropriate randomization methods?
3. Was there concealment of group allocation?
4. For non-randomized trials, did the study employ any other methods to enhance group comparability?
5. Was the process of assigning patients to groups made independently from physician and patient preference?
6. Did the patients in different study groups have similar levels of performance on the outcome of interest at the time they were assigned to groups?
7. Were the study groups comparable for all other important factors at the time they were assigned to groups?
8. Did the study enroll all suitable patients or consecutive suitable patients?
9. Was the comparison of interest prospectively planned?
10. If the patients received ancillary treatment(s), was there a $\leq 5\%$ difference between groups in the proportion of patients receiving each specific ancillary treatment?
11. Were the two groups treated concurrently?
12. Was compliance with treatment $\geq 85\%$ in both of the study's groups?
13. Were patients blinded to the treatment they received?

14. Was the healthcare provider blinded to the groups to which the patients were assigned?
15. Were those who assessed the patient's outcomes blinded to the group to which the patients were assigned?
16. Was the integrity of blinding of patients, physicians, or outcome assessors tested and found to be preserved?
17. Was the outcome measure of interest objective and was it objectively measured?
18. Was a standard instrument used to measure the outcome?
19. Was there $\leq 15\%$ difference in the length of follow up for the two groups?
20. Did $\geq 85\%$ of the patients complete the study?
21. Was there a $\leq 15\%$ difference in completion rates in the study's groups?
22. Was the funding for this study derived from a source that would not benefit financially from results in a particular direction?