Evidence-based Practice Center Systematic Review Protocol

Project Title: Systematic Review of Imaging Tests for the Diagnosis and Staging of Pancreatic Adenocarcinoma

I. Background and Objectives for the Systematic Review

Pancreatic cancer is the fourth most common cause of cancer death among men and women in the United States.[1,2] In 2013, about 45,000 people in the United States will receive a diagnosis of pancreatic cancer, and 38,000 will die of the disease.[3] The median age at diagnosis is 71 years, the overall 5-year survival is 5.8 percent, and the overall age-adjusted mortality rate is 10.8 per 100,000 people per year.[4,5] The most common type of pancreatic cancer is adenocarcinoma (approximately 90% of all pancreatic malignancies).[2] Based on rates from 2007 to 2009, the lifetime risk of receiving a diagnosis of pancreatic cancer is 1.47 percent.[5]

Risk factors for pancreatic cancer include tobacco use; a personal history of pancreatitis, diabetes, or obesity; and a family history of pancreatic cancer.[1] About 10 percent of patients with pancreatic cancer have a positive family history for the disease.[4] Pancreatic cancer incidence rates were reportedly highest among African-American men (21.3 per 100,000) and women (17.6 per 100,000) during 2004 and 2008.[1] The second highest rates were reported for white men (16.8 per 100,000) and women (12.8 per 100,000).[1] The differences between these populations and burden of disease may be related to higher rates of cigarette smoking and diabetes mellitus among African-American men versus white men and higher body mass indices among African-American women versus white women.[1]

Screening for pancreatic adenocarcinoma is not recommended for the general population (e.g., the U.S. Preventive Services Task Force gives it a D recommendation).[6] However, some professional organizations recommend screening those who are at high risk of developing pancreatic cancer. One report[7] suggested that having two or more first-degree relatives with pancreatic cancer is sufficient justification for considering a screening test (or three or more blood relatives with pancreatic cancer, one of whom is a first-degree relative). Further, some genetic risk factors (e.g., Peutz-Jeghers syndrome, Lynch syndrome, and BRCA2, PALB2, and p16 mutations) motivate testing when the patient also has had a first-degree relative with pancreatic cancer.[7]

This review concerns imaging tests to identify and diagnose suspected pancreatic cancer and to determine stage and surgical resectability of the disease.[4,8] The sections below describe the roles of various imaging tests in informing these clinical decisions.

Diagnosis and Staging

Patients often remain asymptomatic or have only nonspecific symptoms such as malaise, fatigue, and loss of appetite until late in the course of the disease—often after it has spread extensively—when weight loss, jaundice, and severe abdominal pain often appear. Due to late

diagnosis, approximately 80 to 85 percent of cases are unresectable (i.e., too advanced to permit surgical resection),[9] and the median survival of patients with unresectable tumors is only 6 to10 months.[10]

Common symptoms leading to suspicion of pancreatic cancer are jaundice, epigastric pain, and weight loss.[11] Signs and symptoms alone, however, are insufficient to diagnose pancreatic cancer. In one study of 70 patients suspected of having pancreatic cancer, only 30 actually had the disease; of the other 40, 16 had irritable bowel syndrome, 9 had other types of intra-abdominal cancer, 8 had pancreatitis, and 7 had other conditions.[12] Thus, additional clinical information—including imaging tests, laboratory values, and biopsies—are important to differentiate these conditions from pancreatic cancer.

Multidetector computed tomography (MDCT) scan is often the first imaging test in a patient whose symptoms suggest pancreatic adenocarcinoma. It provides three-dimensional (3D) multiplanar reconstruction images that enable determination of tumor size, extent, and spread with a standardized pancreas protocol.[13,14] The test does not always differentiate malignant from benign pancreatic lesions, and its ability to detect small tumors or small hepatic/peritoneal metastases is limited. A concern about MDCT is that the procedure exposes the patient to radiation and therefore may increase cancer risk. Also, the quality of the CT protocol, as well as the experience and expertise of the radiologist reading the CT, may influence the accuracy of MDCT for diagnosis and staging of pancreatic adenocarcinoma.

One notable type of MDCT is MDCT with angiography with or without 3D reconstruction.[15] This technology permits more precise imaging of blood vessels than other forms of MDCT. Given its importance, in our review we will separate it from other forms of MDCT.
Other procedures and imaging tests are also used to aid diagnosis of pancreatic adenocarcinoma, including endoscopic ultrasound with fine-needle aspiration (EUS-FNA), positron emission tomography–computed tomography (PET-CT), and magnetic resonance imaging (MRI).

For EUS-FNA, a specialized ultrasound probe is introduced orally and advanced via an endoscope through the upper gastrointestinal tract toward the pancreas. The probe's proximity to the pancreas allows the ultrasound to access and image the entire pancreas, the related vasculature, and associated lymph nodes. The endoscopist can take a small aspiration (FNA) of any suspicious lesions, permitting cytologic evaluation. If the biopsy is adequate, EUS-FNA can distinguish benign from malignant lesions and characterize certain types of lesions (e.g., cystic pancreatic lesions).[16] Reported disadvantages of EUS-FNA include the procedure's invasiveness, dependence on the skill of the endoscopist, and its inability to evaluate distant metastases.[14] The relative newness of EUS-FNA could mean large variation in endoscopists' technical skills. Potential patient harms related to EUS-FNA include perforation and bleeding, pancreatitis, and adverse effects related to sedation.

PET is a whole-body scan whose image highlights where a radioisotope tracer concentrates and is, therefore, particularly useful for detecting distant metastases. The most commonly used radioisotope tracer is fluorodeoxyglucose ^{18}F (FDG). FDG-PET can locate sites metabolically active, such as malignant tumors or sites with inflammation, and may help distinguish malignant tumors from benign pancreatic cysts or other masses that are not metabolically active. FDG-PET and CT can be combined to add precise anatomic localization (from CT) to functional data (from PET). The two scans are acquired concurrently, and the data from each are merged.

MRI is an alternative to MDCT as an initial imaging test for patients with a clinical suspicion of pancreatic adenocarcinoma or as a tool to evaluate the extent of disease. During an MRI

procedure, electromagnetic fields and radiofrequency radiation translate hydrogen nuclei distribution in body tissues into images of anatomic structure. Similar to MDCT, a standardized pancreas protocol is available. MRI may be helpful when characterizing small (<1 cm) hepatic lesions, differentiating an inflammatory pancreatic mass from pancreatic adenocarcinoma, or detecting metastases to the liver.[14] MRI can also be used as an adjunct to CT to better detect extrahepatic disease.[17,18]

As described above, the various available imaging modalities in the diagnosis and staging of pancreatic adenocarcinoma have different strengths and potential benefits and weaknesses and potential harms. At present, there does not appear to be a universal standard delineating which imaging modalities to be used in which cases. This could be due in part to the difficulty of diagnosing and managing such an aggressive cancer, as well as limitations in the relevant evidence. It may also be related to the relative newness of some technologies (e.g., PET-CT). For the clinical team involved in the process, determining which modality or modalities to use, in which clinical situations, and in what order can be difficult.

For the patient, given the poor prognosis of most cases, these differences in modalities and the consequences of their use are important to understand. In addition, elucidating patients' experience with and tolerance of various imaging modalities may help future patients weigh the benefits and harms of the tests and allow them to incorporate their values and priorities. Once pancreatic adenocarcinoma is diagnosed, the stage of disease is a key determinant of clinical management, as well as a key predictor of survival. Most cases are diagnosed at an advanced stage, precluding surgical resection.[1] For localized disease, the 5-year survival is approximately 22 percent.[1] When pancreatic adenocarcinoma is diagnosed at an advanced stage, the 5-year survival is approximately 2 percent.[1]

The most commonly used system for staging pancreatic adenocarcinoma[19] is the 2010 American Joint Committee on Cancer (AJCC) system[20]:

- Stage 0: carcinoma in situ, with neither lymph node involvement nor metastases
- Stage IA: a ≤ 2-cm tumor limited to the pancreas, with neither lymph node involvement nor metastases
- Stage IB: a > 2-cm tumor limited to the pancreas, with neither lymph node involvement nor metastases
- Stage IIA: any size tumor that extends beyond the pancreas but does not involve either the celiac axis or the superior mesenteric artery (SMA) and is without lymph node involvement or metastases
- Stage IIB: the same as IIA, except the lymph nodes are involved
- Stage III: any size tumor that involves the celiac axis or SMA and lymph nodes but without metastases
- Stage IV: any size tumor, any lymph node involvement, and metastases

We note that an exact staging process before surgery (i.e., assigning the patient to stage I/II/III/IV) for pancreatic adenocarcinoma may not be carried out, and the patient's disease is often staged at surgery. For unresectable cases, however, a biopsy is taken and a formal stage is determined to guide the planning of treatments such as chemotherapy.

Resectability

Surgical resection offers the only hope of cure and is decided via multidisciplinary consultation (e.g., surgeon, gastroenterologist, radiologist, oncologist, radiation oncologist). The two key factors in assessing resectability are distant metastasis (which usually indicate unresectability) and blood vessel involvement (which sometimes indicates unresectability, depending on the degree of involvement). The major blood vessels of focus are the superior mesenteric vein (SMV), the portal vein, the celiac artery, the common hepatic artery, and the SMA. According to the 2012 guideline from the National Comprehensive Cancer Network (NCCN) on pancreatic adenocarcinoma[16]:

- A resectable tumor shows no involvement of either the SMV or portal vein and shows "clear fat planes" around the celiac axis, hepatic artery, and SMA, and there are no distant metastases.
- An unresectable tumor has a > 180-degree encasement of the SMA or any celiac abutment, or an unreconstructible SMV/portal vein occlusion, or any aortic invasion/encasement, or any distant metastases
- A "borderline" resectable tumor fits neither of the two categories described above (e.g., some abutment of the SMV/portal vein, a < 180-degree abutment of the SMA). For these cases, the NCCN recommends biopsy and possible neoadjuvant chemotherapy, which may shrink the tumor and permit subsequent resection.

These criteria continue to evolve, as surgical techniques advance and more tumors are deemed resectable via reconstruction of blood vessels.[19]

Regarding the interface between stage and resectability, the AJCC and other professional organizations state that Stages I and II are resectable, but Stages III and IV are not.[20,21] However, others believe that minor arterial involvement (Stage III) may still permit resection.[19,22] Vincent et al.[22] argued that some Stage III cases are borderline resectable and may be appropriate targets for neoadjuvant therapy followed by resection.

A. Rationale for the Review

Pancreatic adenocarcinoma is fatal if untreated, so it is critical to choose the right imaging test and initiate therapy in a timely manner. A comparative effectiveness review (CER) on this topic can assist medical decisions in several ways. First, different imaging tests are believed to have utility in different circumstances (e.g., when suspicious of metastatic disease vs. localized disease) and a clear delineation of the relevant evidence would help guide clinicians and patients in choosing the most appropriate imaging test. Second, the evidence may favor some tests over others and, if so, resources can be devoted to the better tests. Third, it is important to clarify the practice of using a second imaging test: delineating the circumstances under which to order a second test and which test to order; and if a test is ordered, what is its influence on diagnosis, staging, survival, and quality of life. Fourth, the accuracy of any imaging test depends on the operator's and reader's skills and the environment in which the test is performed (e.g., high-volume vs. low-volume centers). Determining the extent to which this is important for various tests, can also help better guide clinicians and patients in the workup process. Fifth, harms are

always a concern; by estimating the actual rates of the various harms of different imaging tests, a CER can help discriminate reasonable fears from unreasonable ones.

II. Key Questions

Question 1

What is the comparative effectiveness of imaging techniques (e.g., MDCT angiography ± 3D reconstruction, other MDCT, EUS-FNA, PET-CT, MRI) for diagnosis of pancreatic adenocarcinoma in adults with suspicious symptoms?

a. What is the accuracy of each imaging technique for diagnosis and assessment of resectability?
b. What is the comparative accuracy of the different imaging techniques for diagnosis and assessment of resectability?
c. What is the comparative diagnostic accuracy of using a single imaging technique versus using multiple imaging techniques?
d. How is test experience (e.g., operative experience, assessor experience, center's annual volume) related to comparative diagnostic accuracy of the different imaging strategies?
e. How are patient factors and tumor characteristics related to the comparative diagnostic accuracy of the different imaging strategies?
f. What is the comparative clinical management after the different imaging strategies when used for diagnosis?
g. What is the comparative impact of the different imaging strategies on long-term survival and quality of life when used for diagnosis?

Question 2

What is the comparative effectiveness of imaging techniques (e.g., MDCT angiography ± 3D reconstruction, other MDCT, EUS-FNA, PET-CT, MRI) for *staging* of pancreatic adenocarcinoma among adults with a diagnosis of pancreatic adenocarcinoma?

a. What is the staging accuracy of each imaging technique (for tumor size, lymph node status, vessel involvement, metastases, stage [I–IV], and resectability)?
b. What is the comparative staging accuracy among the different imaging techniques?
c. What is the comparative staging accuracy of using a single imaging technique versus using multiple imaging techniques?
d. How is test experience (e.g., operative experience, assessor experience, center's annual volume) related to comparative staging accuracy of the different imaging strategies?
e. How are patient factors and tumor characteristics related to the comparative staging accuracy of the different imaging strategies?
f. What is the comparative clinical management of the different imaging strategies when used for staging?
g. What is the comparative impact of the different imaging strategies on long-term survival and quality of life when used for staging?

Question 3

What are the rates of harms of imaging techniques (e.g., MDCT angiography ± 3D reconstruction, other MDCT, EUS-FNA, PET-CT, MRI) when used to diagnose and/or stage pancreatic adenocarcinoma?

a. How are patient factors related to the harms of different imaging techniques?
b. What are patient perspectives on the tolerance of different imaging techniques and the balance of benefits and harms of different imaging techniques?

Question 4

What is the comparative screening accuracy of imaging techniques (e.g., MDCT angiography ± 3D reconstruction, other MDCT, EUS-FNA, PET-CT, MRI) in high-risk asymptomatic adults (i.e., those at genetic or familial risk of pancreatic adenocarcinoma)?

Many Key Questions (KQs) involve an assessment of "accuracy." The best way to determine the accuracy of an imaging test is to compare its interpretation to a *gold standard*. In this context, the only gold standard is surgical findings. However, most patients do not undergo surgical resection because their tumors are unresectable; as such, a gold standard is not available for most patients. Instead, the *imperfect reference standard* is clinical followup. Thus, the imaging results are compared against a gold standard for patients who undergo resection and an imperfect reference standard for patients who do not.

The draft KQs were posted for public comment between February 26 and March 25, 2013, on the Effective Health Care Program Web site. No comments were received. However, based on discussion among Evidence-based Practice Center (EPC) team members, we made five changes to the KQs (these changes are all reflected in the list of questions above):

- We added the aspect of resectability to KQ 1 (diagnosis) because some patients are not formally staged before surgery.
- We added the precise stage (stage IA, IB, IIA, IIB, III, or IV) to KQ 2 (staging) because some studies may report accuracy data based on stage (e.g., percentage of tumors judged to be stage IB by imaging and were later found to be stage IB during surgery).
- We changed KQ 3 (harms) to "rates of harms" rather than simply "harms" to clarify the focus on how common the harms are, not simply what the possible harms are.
- We added a subquestion to KQ 3 regarding the association between patient factors and the harms of different imaging techniques.
- We added a subquestion to KQ 3 regarding patient perspectives on tolerance of different imaging techniques, as well as the balance of benefits and harms.

The table below summarizes the PICO (**p**opulation, **i**nterventions, **c**omparators, and **o**utcomes) for each KQ. The only timing (**t**) issue concerns the outcome of long-term survival and quality of life, for which we will define "long-term" as 1 year or more. All settings (**s**) are potentially relevant. In the table, population P1 is symptomatic patients being assessed for possible pancreatic adenocarcinoma; population P2 is adults with known pancreatic adenocarcinoma;

population P3 is asymptomatic adults at high risk for pancreatic adenocarcinoma. High risk encompasses those with either a genetic or familial risks such as having two or more first-degree relatives with pancreatic cancer; three or more blood relatives with pancreatic cancer, one of whom is a first-degree relative; or having Peutz-Jeghers syndrome, Lynch syndrome, or a BRCA2, PALB2, or p16 mutation in addition to having a first-degree relative with pancreatic cancer.

KQ	Population	Interventions	Comparators	Outcomes
1a	P1	MDCT angiography ± 3D reconstruction, other MDCT, EUS-FNA, PET-CT, or MRI	None	• Diagnostic accuracy as determined by surgical findings and/or clinical followup • Accuracy of resectability judgment
1b	P1	Same list as KQ 1a	Another test from the list of interventions for KQ 1a	• Diagnostic accuracy as determined by surgical findings and/or clinical followup • Accuracy of resectability judgment
1c	P1	Single imaging test: same list as KQ 1a	Multiple tests from the list of interventions for KQ 1a	Diagnostic accuracy as determined by surgical findings and/or clinical followup
1d	P1	One test: most vs. least experience	Another test: most vs. least experience	Diagnostic accuracy as determined by surgical findings and/or clinical followup
1e	P1	Patient factors or tumor characteristics	Comparator patient factor (e.g., age) or tumor characteristic (e.g., head or tail of pancreas)	Diagnostic accuracy as determined by surgical findings and/or clinical followup
1f	P1	Same list as KQ 1a	Same as list of interventions for KQ 1a	Clinical management (e.g., the percentage of patients in whom resection is attempted)
1g	P1	Same list as IQ 1a	Same as list of interventions for KQ 1a	• Overall survival (minimum 1-year followup) • Pancreatic adenocarcinoma-specific survival (minimum 1-year follow-up) • Quality of life (e.g., SF-36®; minimum 1-year follow-up)
2a	P2	Same list as KQ 1a	None	Staging accuracy as determined by surgical findings and/or clinical followup: • T stage • N stage • M stage • Stage I, II, III, or IV • Vessel involvement • Resectability
2b	P2	Same list as KQ 1a	Same list	Staging accuracy as determined by surgical findings and/or clinical followup (same list as KQ 2a above)
2c	P2	Single imaging test: same list as KQ 1a	Multiple imaging tests: same as list of interventions for KQ 1a	Staging accuracy as determined by surgical findings and/or clinical followup (same list as KQ 2a above)

KQ	Population	Interventions	Comparators	Outcomes
2d	P2	One test: greatest vs. least experience	Another test: greatest vs. least experience	Staging accuracy as determined by surgical findings and/or clinical followup (same list as KQ 2aabove)
2e	P2	One test: effect of patient factor or tumor characteristic	Another test: Effect of patient factor or tumor characteristic	Staging accuracy as determined by surgical findings and/or clinical followup (same list as above)
2f	P2	Same list as KQ 1a	Same as list of interventions for KQ 1a	Clinical management (e.g., the percentage of patients in whom resection is attempted)
2g	P2	Same list as KQ 1a	Same as list of interventions for KQ 1a	• Overall survival (minimum 1-year followup) • Pancreatic adenocarcinoma-specific survival (minimum 1-year followup) • Quality of life (e.g., SF-36®) (minimum 1- year followup)
3	P1 or P2 or P3	Same list as KQ 1a	None	• Radiation from MDCT (e.g., carcinogenic effects) • Adverse reactions to contrast agents • Adverse reaction to radiopharmaceuticals • Pancreatitis from EUS-FNA • Perforation or bleeding from EUS-FNA • Sedation-related effects of EUS-FNA (e.g., nausea, vomiting)
3a	P1 or P2 or P3	Patient factor	Comparator patient factor	See the list of harms for KQ 3
3b	P1 or P2 or P3	Same list as KQ 1a	Any	Patient perspectives on imaging techniques, including tolerance, satisfaction, preference, and balance of benefits and harms
4	P3	Same list as KQ 1a	Same as list of interventions for KQ 1a	Screening accuracy as determined by surgical findings and/or clinical followup

Abbreviations: EUS-FNA = endoscopic ultrasound with fine needle aspiration; KQ = key question; MDCT = multidetector computed tomography; PET-CT = combined position emission tomography and computed tomography; MRI = magnetic resonance imaging; SF-36 = Short Form (36) Health Survey

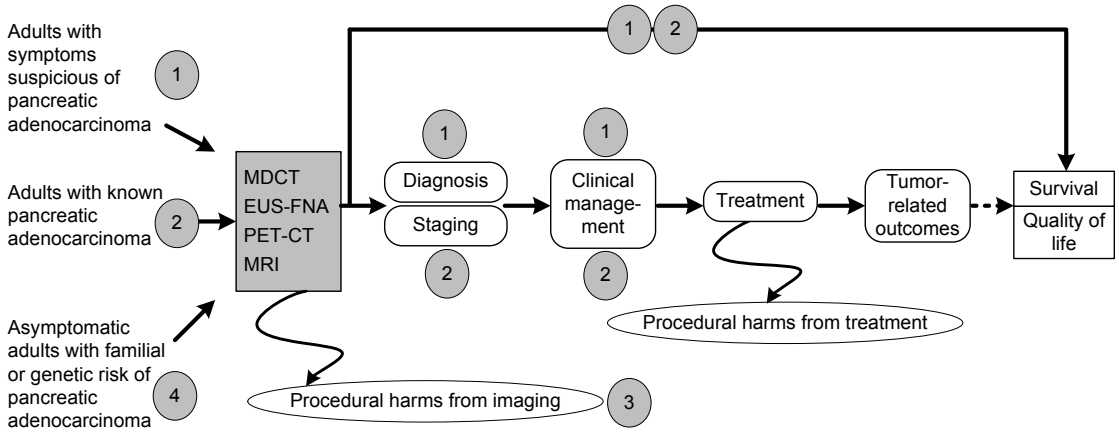
III. Analytic Framework

An analytic framework illustrating the connections between the populations of interest, the imaging techniques, and the outcomes is shown in Figure 1 below. Populations that are undergoing or have undergone treatment for pancreatic adenocarcinoma are outside the scope.

The populations of interest enter the diagram at the left, undergo diagnosis (KQ 1), staging (KQ 2), and then commence treatment. Some outcomes such as test performance can be measured immediately after performing the tests, but the most important outcomes (such as long-term survival and quality of life) are measured after completion of treatment.

An important factor in selecting an imaging modality is the availability and accessibility of that modality. Although this factor will not be addressed formally in the review as a KQ, we plan to collect and provide relevant information about the availability and accessibility of imaging modalities and information about current patterns of care as available. This information will be presented in the background and discussion sections to help place the evidence review findings in context.

Figure 1. Analytic Framework



Note: Circled numbers, e.g., (1) denote Key Questions addressed by the systematic review. MDCT – Multidimensional computed tomography; EUS-FNA – Endoscopic ultrasound with fine needle aspiration; PET-CT – Simultaneous positron emission tomography and computed tomography; MRI – Magnetic resonance imaging

Figure 1. This figure depicts the Key Questions (circled numbers) within the context of the PICOTS (**p**opulations, **i**nterventions, **c**omparators, **o**utcomes, **t**ime points, and **s**etting). In general, the figure illustrates how different types of patients (the three populations listed on the left) can undergo different imaging tests (listed in the large box), resulting in the intermediate outcomes of diagnostic accuracy, staging accuracy, and clinical management decisions. Treatment is intended to improve (if possible) the patient-oriented outcomes listed to the right: survival and quality of life. Also, procedural harms of the imaging procedures may occur.

IV. Methods

A. Criteria for Inclusion/Exclusion of Studies in the Review

Our criteria are listed in five categories below: 1) publication criteria, 2) study design criteria, 3) patient criteria, 4) test criteria, and 5) data criteria. After we discuss the five categories, we provide a table summarizing all of the inclusion/exclusion criteria.

Publication Criteria

1. **Full-length articles.** The article must be published as a full-length peer-reviewed study. Abstracts and meeting presentations will not be included because they do not include

sufficient details about experimental methods to permit an evaluation of study design and conduct and because they may also contain only a subset of measured outcomes.[23,24] In addition, it is not uncommon for abstracts that are published as part of conference proceedings to have inconsistencies when compared with the final publication of the study or to describe studies that are never published as full articles.[25-29]

2. **Redundancy.** To avoid double-counting of patients when several reports of the same or overlapping groups of patients are 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 reports an included outcome that was not reported by other publications of that study or (b) a publication with lower patient enrollment reports longer followup data for an outcome.

3. **English language.** Moher et al. have demonstrated that exclusion of non–English-language studies from meta-analyses has little impact on the conclusions drawn.[30] Juni et al. found that non–English-language studies typically were of higher risk of bias and that excluding them had little effect on effect-size estimates in the majority of the meta-analyses they examined.[31] Although we recognize that in some situations exclusion of non–English-language 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 translation of studies.

4. **Publication date.** We will include studies published since January 1, 2000. Older articles likely included outdated technologies. For studies of harms of imaging technologies that did not specifically involve pancreatic adenocarcinoma (i.e., any clinical indication), they must have been published since January 1, 2009.

Study Design Criteria

1. **Key Questions on single-test accuracy.** For KQs 1a and 1b, which address the performance of a single imaging test against a reference standard, we will first attempt to answer the question using systematic reviews only. EPC guidance by White et al.[32] states that existing systematic reviews can replace de novo processes in CERs if the reviews are of sufficient *relevance* and *quality*. For relevance, we will refer to the PICOTS-SD for the pertinent subquestion, and these seven components—**p**opulations, **i**nterventions, **c**omparisons, **o**utcomes, **t**ime points, **s**etting, and **s**tudy **d**esign—will be the seven inclusion criteria. For sufficient quality, see section D below on risk of bias. If the included systematic reviews are insufficient to answer either question on single-test accuracy, we will include primary studies of single-test accuracy. For such primary studies, we will consult with the Technical Expert Panel (TEP) to determine suitable inclusion criteria, given the time frame of the project.

2. **Any Key Questions comparing two or more tests.** The study must have compared both tests to a reference standard.

3. **Any Key Questions on single versus multiple tests, test experience, patient factors (e.g., age), or tumor characteristics (e.g., head or tail of pancreas).** The study must have made a

comparison of data to address the question. For example, for test experience, the difference between MDCT and EUS-FNA may depend on the experience of the imaging centers (e.g., higher volume centers may find less of a difference in these technologies than lower-volume centers).

4. **Any Key Questions involving comparative clinical management or long-term survival/quality of life.** Some patients must have received one of the imaging tests, and a separate group of patients must have received a different imaging test. This design permits a comparison of how the choice of test may influence management and/or survival and/or quality of life.

5. **Key Question 3 on the rates of procedural harms.** We will include any reported data on more than 100 patients on the harms of imaging procedures from studies already included for KQs 1 and 2, and we will also include single-test studies enrolling at least 100 patients that reported data on the harms of an imaging procedure. In addition, we will search specifically for reports of harms and adverse events associated with the use of each specific imaging modality (such as radiation exposure and reactions to contrast agents) that were published in 2009 or later.

6. **Key Question 3b on patient perspectives of imaging tests.** Any study design will be accepted.

Patient Criteria

1. To be included, the study must have reported data obtained from groups of patients in which at least 85 percent of the patients were from one of the patient populations of interest. If a study reported multiple populations, it must have reported data separately for one or more of the populations of interest.

2. At least 85 percent of patients must have been age 18 or older, or data must have been reported separately for patients age 18 or older.

3. Studies on the screening, diagnosis, and staging *primary* pancreatic adenocarcinoma will be included. Testing for *recurrent* pancreatic cancer is excluded.
4. Data on imaging tests performed after any form of treatment (e.g., neoadjuvant chemotherapy) will be excluded, but pretreatment imaging data will be considered.

Test Criteria

1. Only studies of the imaging tests of interest will be included (listed in the Key Questions section above).

2. Studies of CT that did not explicitly state that CT was MDCT will be assumed to be MDCT. Given our publication date criterion of 2000 and later, we believe it safe to assume that CT performed in such studies was MDCT.

Data Criteria

1. The study must report data pertaining to one of the outcomes of interest (see the Key Questions section above).

    a. For accuracy outcomes (KQ 1a–e, KQ 2a–e, and KQ 4), this means reporting enough information for one to calculate both sensitivity and specificity, along with corresponding confidence intervals.
    b. For clinical management (KQ 1f and KQ 2f), this means reporting the percentage of patients who received a specific management strategy after undergoing each imaging test (a separate group of patients corresponding to each imaging test).
    c. For long-term survival (KQ 1g and KQ 2g), this means either reporting median survival after each imaging test (separate groups of patients), or mortality rates at a given time point (separate groups of patients), or another patient survival measure such as a hazard ratio.
    d. For quality of life (KQ 1g and KQ 2g), this means reporting data on a previously tested quality-of-life instrument (such as the Short Form [36] Health Survey) after each imaging test (separate groups of patients).
    e. For harms (KQ 3), there is a statement in the Methods section that harms/complications would be measured, reporting the occurrence of a procedure-related harm and the ability to determine the rate of that harm or reporting that no harms or complications occurred as a result of the procedure.
    f. For patient perspectives (KQ 3b), this means reporting the responses patients gave about their opinions or experience after having undergone one or more of the imaging tests.

2. Regarding the minimum patient enrollment, for studies comparing imaging tests (KQ 1b–g and KQ 2b–g), we required data on at least 10 patients per imaging test. We also used a minimum of 10 for KQ 3b on patient perspectives of imaging tests. For single-test studies of accuracy (KQ 1a and KQ 2a), if systematic reviews cannot answer the KQ, we required data on at least 100 patients. We also used a minimum of 100 patients for data on harms (KQ 3) or screening (KQ 4).

3. For all KQs, the reported data must have included at least 50 percent of the patients who had initially enrolled in the study.

4. Studies that report data by tumor (e.g., percentage of pancreatic adenocarcinoma *tumors* that were correctly detected) instead of by patient (e.g., percentage of enrolled *patients* that were correctly given a diagnosis of pancreatic adenocarcinoma) will not be excluded for this reason. However, the tumor-based data will be separated from the patient-based data because they measure different types of accuracy.

The table on the next pages summarizes all of the inclusion criteria discussed above.

	KQ 1a & KQ 2a (single-test accuracy)	KQ 1b, KQ 1c, KQ 2b, & KQ 2c (comparative accuracy)	KQ 1d, KQ 1e, KQ 2d, & KQ 2e (influence of various factors on comparative accuracy)	KQ 1f & KQ 2f (comparative patient management)	KQ 1g & KQ 2g (comparative survival/quality of life)	KQ 3 & KQ 3a (rates of harms, patient factors)	KQ 3b (patient perspectives on harms/ convenience)	KQ 4 (screening)
Inclusion criteria for all KQs	• Full-length articles • English language • No double-counting of patients across publications • Reported data on one or more of the five imaging tests of interest (MDCT angiography, other MDCT, EUS-FNA, PET-CT, and MRI) • Images taken before any treatment • Clinical purpose for primary pancreatic adenocarcinoma screening, diagnosis, and staging rather than for recurrence (except for the harms question in which we included data on other clinical conditions) • At least 85 percent of patients were in the relevant population (or data for them was reported separately) • At least 85 percent adults (or adult data reported separately) • Reported data on at least 50 percent of those enrolled							
Publication date	2000+					• 2000+ for studies involving screening, diagnosis, and staging of pancreatic adeno-carcinoma • 2009+ for other studies	2000+	

	KQ 1a & KQ 2a (single-test accuracy)	KQ 1b, KQ 1c, KQ 2b, & KQ 2c (comparative accuracy)	KQ 1d, KQ 1e, KQ 2d, & KQ 2e (influence of various factors on comparative accuracy)	KQ 1f & KQ 2f (comparative patient management)	KQ 1g & KQ 2g (comparative survival/quality of life)	KQ 3 & KQ 3a (rates of harms, patient factors)	KQ 3b (patient perspectives on harms/ convenience)	KQ 4 (screening)
Study design	Systematic review: If those do not answer the question, then study must have included patients with different true statuses	Included patients with different true statuses and made a relevant comparison		Separate groups of patients received different imaging tests		Any		Included patients with different true statuses
Minimum N at followup	• Not applicable if systematic reviews are adequate • If not adequate, then 100	10				100	10	100

	KQ 1a & KQ 2a (single-test accuracy)	KQ 1b, KQ 1c, KQ 2b, & KQ 2c (comparative accuracy)	KQ 1d, KQ 1e, KQ 2d, & KQ 2e (influence of various factors on comparative accuracy)	KQ 1f & KQ 2f (comparative patient management)	KQ 1g & KQ 2g (comparative survival/quality of life)	KQ 3 & KQ 3a (rates of harms, patient factors)	KQ 3b (patient perspectives on harms/ convenience)	KQ 4 (screening)
Data	• Systematic review estimates of sensitivity, specificity, and CIs • If reviews not adequate, then studies must have had sufficient information to calculate sensitivity, specificity, and CIs	Sufficient information to calculate sensitivity, specificity, and CIs separately for ≥ 2 tests	Sufficient information to calculate sensitivity, specificity, and CIs separately for different factors	Comparative rates of receiving specific form of management	Comparative survival data or comparative quality-of-life data using a previously tested instrument	Rates of harms, or a statement that no harms or complications occurred	Responses of patients asked about their opinion of or experience with an imaging test they received	Sufficient information to calculate sensitivity, specificity, and their CIs

Abbreviations: CI = confidence interval; KQ = key questions

B. 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 EPC Information Center; searches will follow established systematic review protocols. 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, the Internet, Medscape, National Guideline Clearinghouse™, and the U.S. Food and Drug Administration. We conducted separate searches for systematic reviews on imaging for diagnosis/staging/screening/harms in the context of pancreatic adenocarcinoma, primary studies on one or more of these topics, and primary studies of the harms of imaging procedures regardless of the clinical purpose. This last search was conducted because several imaging tests have harms that are essentially the same regardless of clinical indication (e.g., radiation dose after whole-body PET-CT). A sample search strategy is shown in Appendix A.

Literature screening will be performed in duplicate using the Distiller SR database (Evidence Partners, Ottawa, Canada). Literature search results will initially be screened for relevancy. Relevant abstracts will be screened against the study inclusion criteria in duplicate. Studies that appear to meet the study inclusion criteria according to either screener will be retrieved in full and screened again in duplicate against the study inclusion criteria. All disagreements at the full article level will be resolved by consensus discussion among the two original screeners, with adjudication if necessary by a senior reviewer.

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

C. Data Abstraction and Data Management

Data will be abstracted using Microsoft Excel. Duplicate abstraction on a 10-percent random sample will be used to ensure accuracy. All discrepancies will be resolved by consensus discussion among the two original abstracters and an additional third person. Elements to be abstracted include general study characteristics (e.g., country, setting, study design, enrolled N), patient characteristics (e.g., age, sex, comorbidities), details of the imaging methodology (e.g., radiotracer, timing of test), risk of bias items, and outcomes data.

D. Assessment of Methodological Risk of Bias of Individual Studies

For systematic reviews of single-test accuracy, EPC guidance by White et al.[32] suggests that EPCs assess the quality of an existing systematic review by using a revised AMSTAR (Assessment of Multiple Systematic Reviews) instrument. The items we plan to use for this appear in Appendix B. For each included review, two analysts will independently answer 11 items and independently assign the review as having either low risk of bias or moderate/high risk of bias. Discrepancies in the category assignment will be resolved by consensus or by a third party if consensus cannot be reached. Reviews at moderate/high risk of bias will be excluded.

For studies of single-test performance, if systematic reviews are insufficient to answer questions of single-test accuracy, we will use an internal validity rating scale for diagnostic studies to assess the risk of bias of each individual study. This instrument is based on a modification of the QUADAS instrument with reference to empirical studies of design-related bias in diagnostic test studies (see Appendix B).[33-35] Each question in the instrument addresses an aspect of study design or conduct that can help protect against bias, such as enrolling consecutive or a random sampling of patients or blinding image readers to clinical information about the patient. 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 comparing two or more diagnostic tests, we will consider the items on the modified QUADAS-2, as well as four additional items that specifically address bias in the comparison of diagnostic tests (see Appendix B).

For studies addressing clinical outcomes, 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 the ECRI Institute[36] with respect to the impact of study design on bias in comparative studies and is consistent with the guidance in the *Methods Guide for Effectiveness and Comparative Effectiveness Reviews* (hereafter *Methods Guide*; see Appendix B).[37] 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.

Studies will be rated as having "low," "medium," or "high" risk of bias. The 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 having "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.
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).

E. Data Synthesis

Specifically for questions addressing individual test performance (accuracy), we will draw evidence from previous systematic reviews as available and will consult the primary literature only as necessary to update or supplement the information available in previous systematic reviews. As the *Methods Guide* recommends, we will summarize all relevant, high-quality reviews.[32]

For studies of single-test performance, if systematic reviews are insufficient to answer the KQ, we will meta-analyze the data reported by the studies using a bivariate mixed-effects binomial regression model as described by Harbord et al.[38] All such analyses are computed by the STATA® 10.0 statistical software package using the "midas" command.[39] 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 Meta-Disc software package.[40]

For questions comparing imaging tests, we will synthesize the evidence from the primary studies themselves. For statistical analyses comparing the accuracy of two imaging tests, we will use the EPC methods described by Trikalinos et al.[41] 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 studies of clinical outcomes, we will compute effect sizes and measures of variance using standard methods and will perform a DerSimonian and Laird random-effects meta-analysis using Comprehensive Meta-Analysis (CMA) software (Biostat, Inc., Englewood, NJ). Meta-regression and subgroup analysis will be used to explore possible causes of heterogeneity. Potential covariates include population descriptors, tumor size and type, and country and setting of care.

F. Grading the Strength of Evidence for Individual Comparisons and Outcomes

We will use the EPC system for grading evidence on diagnostic tests as described in the EPC guidance chapter by Singh et al.[42] This system uses up to eight domains as input (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 rating of the strength of evidence: high, moderate, low, or insufficient. This rating is made separately for each outcome of each comparison of each KQ.

A rating of insufficient will be given when the evidence does not permit a conclusion for the outcome of interest for that KQ. For single-test accuracy, this means whether the evidence permits an estimate of the sensitivity and specificity of the test. For comparative test accuracy, it means whether the evidence permits a conclusion that either (1) evidence indicates that test A is more accurate, (2) evidence indicates that test B is more accuracy, or 3) evidence indicates that tests A and B are similarly accurate. If none of these conclusions can be drawn, then evidence is deemed insufficient for questions on comparative test accuracy.

If the evidence is sufficient to permit a conclusion, then the rating is deemed high, moderate, or low. The rating 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 input to the rating:

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 rating than a set of studies at moderate or high risk of bias.

Directness. For questions on test accuracy, data on accuracy directly address the question, so those data will be considered direct. For questions on other outcomes (e.g., long-term survival), data on the actual outcomes will be required to be judged direct.

Consistency. For questions on single-test accuracy, consistency will be measured meta-analytically using the tau statistical test. For questions comparing the accuracy of two or

more tests and for other comparative questions, consistency will be judged according to whether the studies' findings suggested the same direction of effect.

Precision. For questions on single-test accuracy, precision will be measured using the random-effects summary confidence interval. For questions comparing the accuracy of two or more tests and for other comparative questions, the evidence will be considered sufficiently precise if the data show a statistically significant difference (between groups or between tests) or if the data demonstrate approximate equivalence by ruling out the minimum important difference.

Publication 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 decrease the strength of the evidence.

Dose-response association. This domain is only relevant with respect to the radiation dose for CT. One possibility is that higher doses result in higher accuracy of CT. If the evidence shows that CT is more accurate than another imaging technique and that the difference is even larger in studies that used higher CT doses, it would generally increase the strength of evidence.

All plausible confounders would reduce the effect. This domain means that a set of studies may be biased *against* finding a difference between two interventions 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 rating of the strength of evidence.

Strength of association. This domain is relevant only to comparative KQs and 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 rating of the strength of evidence.

G. 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, sex) and then assess whether the included studies enrolled similar patients. Some aspects of the 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 based on setting will be analyzed separately.

V. References

1. American Cancer Society. Cancer facts & figures 2012. Atlanta, GA: American Cancer Society (ACS); 2012. Available at http://www.cancer.org/acs/groups/

2. Sharma C, Eltawil KM, Renfrew PD, et al. Advances in diagnosis, treatment and palliation of pancreatic carcinoma: 1990–2010. World J Gastroenterol. 2011 Feb 21;17(7):867-97. PMID: 21412497.
3. National Cancer Institute. Pancreatic Cancer. http://www.cancer.gov/cancertopics/types/pancreatic.
4. Dabizzi E, Assef MS, Raimondo M. Diagnostic management of pancreatic cancer. Cancers. 2011 Mar;3(1):494-509.
5. National Cancer Institute. Surveillance Epidemiology and End Results. SEER Stat Fact Sheets: Pancreas. http://seer.cancer.gov/statfacts/html/pancreas.html.
6. U.S. Preventive Services Task Force. Screening for Pancreatic Cancer: Recommendation Statement. http://www.uspreventiveservicestaskforce.org/3rduspstf/pancreatic/pancrers.htm.
7. Canto MI, Harinck F, Hruban RH, et al. International Cancer of the Pancreas Screening (CAPS) Consortium summit on the management of patients with increased risk for familial pancreatic cancer. Gut. 2013 Mar;62(3):339-47. PMID: 23135763.
8. Dewitt J, Devereaux BM, Lehman GA, et al. Comparison of endoscopic ultrasound and computed tomography for the preoperative evaluation of pancreatic cancer: a systematic review. Clin Gastroenterol Hepatol. 2006 Jun;4(6):717-25; quiz 664. PMID: 16675307.
9. National Cancer Institute. Pancreatic Cancer Treatment (PDQ®): Health Professional Version. http://www.cancer.gov/cancertopics/pdq/treatment/pancreatic/HealthProfessional/page1/AllPages/Print.
10. Benson AB 3rd, Myerson RJ, Sasson AR. Pancreatic, neuroendocrine GI, and adrenal cancers. In: Haller DG, Wagman LD, Camphausen KA, Hoskins WJ, eds. Cancer Management: A Multidisciplinary Approach. 13th ed. New York: UBM Medica LLC; 2011.
11. Fernandez-del Castillo C. Clinical manifestations, diagnosis, and staging of exocrine pancreatic cancer. UpToDate. 2013 Jan 14. Available at http://www.uptodate.com/contents/clinical-manifestations-diagnosis-and-staging-of-exocrine-pancreatic-cancer?source=search_result&search=Clinical+manifestations%2C+diagnosis%2C+and+staging+of+exocrine+pancreatic+cancer&selectedTitle=1%7E134.
12. DiMagno EP, Malagelada JR, Taylor WF, et al. A prospective comparison of current diagnostic tests for pancreatic cancer. N Engl J Med. 1977 Oct 6;297(14):737-42. PMID: 895803.
13. Callery MP, Chang KJ, Fishman EK, et al. Pretreatment assessment of resectable and borderline resectable pancreatic cancer: expert consensus statement. Ann Surg Oncol. 2009 Jul;16(7):1727-33. PMID: 19396496.
14. Wong JC, Lu DS. Staging of pancreatic adenocarcinoma by imaging studies. Clin Gastroenterol Hepatol 2008 Dec;6(12):1301-8. PMID: 18948228.
15. Fang CH, Zhu W, Wang H, et al. A new approach for evaluating the resectability of pancreatic and periampullary neoplasms. Pancreatology. 2012 Jul-Aug;12(4):364-71. PMID: 22898639.
16. National Comprehensive Cancer Network. NCCN Guidelines: Pancreatic Adenocarcinoma. http://www.nccn.org/professionals/physician_gls/f_guidelines.asp.

17. Schima W, Ba-Ssalamah A, Goetzinger P, et al. State-of-the-art magnetic resonance imaging of pancreatic cancer. Top Magn Reson Imaging. 2007 Dec;18(6):421-9. PMID: 18303400.
18. Vachiranubhap B, Kim YH, Balci NC, et al. Magnetic resonance imaging of adenocarcinoma of the pancreas. Top Magn Reson Imaging. 2009 Feb;20(1):3-9. PMID: 19687720.
19. Tamm EP, Balachandran A, Bhosale PR, et al. Imaging of pancreatic adenocarcinoma: update on staging/resectability. Radiol Clin North Am. 2012 May;50(3):407-28. PMID: 22560689.
20. Edge SB, Byrd DR, Compton CC, et al, eds. AJCC Cancer Staging Manual. 7th ed. New York: Springer; 2010.
21. Hidalgo M. Pancreatic cancer. N Engl J Med. 2010 Apr 29;362(17):1605-17. PMID: 20427809.
22. Vincent A, Herman J, Schulick R, et al. Pancreatic cancer. Lancet. 2011 Aug 13;378(9791):607-20. PMID: 21620466.
23. Chalmers I, Adams M, Dickersin K, et al. A cohort study of summary reports of controlled trials. JAMA. 1990 Mar 9;263(10):1401-5. PMID: 2304219.
24. Neinstein LS. A review of Society for Adolescent Medicine abstracts and Journal of Adolescent Health Care articles. J Adolesc Health Care. 1987 Mar;8(2):198-203. PMID: 3818406.
25. Dundar Y, Dodd S, Williamson P, et al. Case study of the comparison of data from conference abstracts and full-text articles in health technology assessment of rapidly evolving technologies: does it make a difference? Int J Technol Assess Health Care. 2006 Summer;22(3):288-94. PMID: 16984055.
26. De Bellefeuille C, Morrison CA, Tannock IF. The fate of abstracts submitted to a cancer meeting: factors which influence presentation and subsequent publication. Ann Oncol. 1992 Mar;3(3):187-91. PMID: 1586615.
27. Scherer RW, Langenberg P, von Elm E. Full publication of results initially presented in abstracts. Cochrane Database Syst Rev. 2007 Apr 18;(2):MR000005. PMID: 17443628.
28. Yentis SM, Campbell FA, Lerman, J. Publication of abstracts presented at anaesthesia meetings. Can J Anaesth. 1993 Jul;40(7):632-4. PMID: 8403137.
29. Marx WF, Cloft HJ, Do HM, et al. The fate of neuroradiologic abstracts presented at national meetings in 1993: rate of subsequent publication in peer-reviewed, indexed journals. Am J Neuroradiol 1999 Jun-Jul;20(6):1173-7. PMID: 10445467.
30. Moher D, Pham B, Klassen TP, et al. What contributions do languages other than English make on the results of meta-analyses? J Clin Epidemiol. 2000 Sep;53(9):964-72. PMID: 11004423.
31. Juni P, Holenstein F, Sterne J, et al. Direction and impact of language bias in meta-analyses of controlled trials: empirical study. Int J Epidemiol. 2002 Feb;31(1):115-23. PMID: 11914306.
32. White CM, Ip S, McPheeters M, et al. Using existing systematic reviews to replace de novo processes in conducting comparative Effectiveness reviews. In: Methods Guide for Effectiveness and Comparative Effectiveness Reviews. AHRQ Publication No. 10(12)-EHC063-EF. Rockville, MD: Agency for Healthcare Research and Quality; 2012:163-78. Available at

http://effectivehealthcare.ahrq.gov/ehc/products/60/318/MethodsGuide_Prepublication-Draft_20120523.pdf.

33. Whiting P, Rutjes AW, Reitsma JB, et al. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol. 2003 Nov 10;3(1):25. PMID: 14606960.
34. Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. JAMA. 1999 Sep 15;282(11):1061-6. PMID: 10493205.
35. Whiting PF, Rutjes AW, Westwood ME, et al; QUADAS-2 Group. 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.
36. Treadwell JR, Tregear SJ, Reston JT, et al. A system for rating the stability and strength of medical evidence. BMC Med Res Methodol. 2006 Oct 19;6:52. PMID: 17052350.
37. Viswanathan M, Ansari MT, Berkman ND, et al. Assessing the risk of bias of individual studies in systematic reviews of health care interventions. In: Methods Guide for Effectiveness and Comparative Effectiveness Reviews. AHRQ Publication No. 10(12)-EHC063-EF. Rockville, MD: Agency for Healthcare Research and Quality; 2012:69-97. Available at http://effectivehealthcare.ahrq.gov/ehc/products/60/318/MethodsGuide_Prepublication-Draft_20120523.pdf.
38. 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.
39. STATA® Data Analysis and Statistical Software. Version 10.0.College Station, TX: StataCorp; 1984-2007.
40. Zamora J, Abraira V, Muriel A, et al. Meta-DiSc: a software for meta-analysis of test accuracy data. BMC Med Res Methodol. 2006 Jul 12;6:31. PMID: 16836745.
41. 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; January 2013. Available at http://www.ncbi.nlm.nih.gov/books/NBK148804/pdf/TOC.pdf.
42. Singh S, Chang S, Matchar DB, et al. Grading a body of evidence on diagnostic tests In: Chang SM, Matchar DB, eds. Methods Guide for Medical Test Reviews. AHRQ Publication No. 12-EHC017. Rockville, MD: Agency for Healthcare Research and Quality; 2012; chapter 7. Available at http://effectivehealthcare.ahrq.gov/ehc/products/246/558/Methods-Guide-for-Medical-Test-Reviews_Full-Guide_20120530.pdf.
43. Shea BJ, Grimshaw JM, Wells GA, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol. 2007 Feb 15;7:10. PMID: 17302989.

VI. Definition of Terms

Blood vessel involvement — The extent to which the tumor surrounds or involves major blood vessels. The degree of surrounding (e.g., <180° or >180°) and the specific blood vessel (e.g.,

superior mesenteric artery) will influence resectability. Venous involvement is generally more amenable to resection than arterial involvement.

M staging — In the American Joint Committee on Cancer (AJCC) TNM system, M0 denotes a primary tumor that has not metastasized, and M1 denotes metastases. Metastases are unresectable.

Metastases – Spread of the primary pancreatic tumor to other distant parts of the body (e.g., liver, peritoneum)

N staging — In the AJCC TNM system, N0 means no regional lymph node involvement, and N1 means the regional lymph nodes are involved. Lymph node involvement does not play a critical role in determining resectability.

Pancreatic adenocarcinoma — The most common type of pancreatic cancer that is a solid tumor.

Pancreatitis — Inflammation of the pancreas.

Radiation — A harm of computed tomography that can increase the risk of developing cancer.

Resectability — The degree to which the tumor can be safely removed surgically. Resection is the only chance of cure for patients who have pancreatic adenocarcinoma.

Sensitivity — The performance or likelihood of an imaging test to correctly detect cancer. It is computed by dividing the number of patients who test positive on the imaging test by the number of patients who were actually positive via the gold standard test.

Specificity — The performance or likelihood of an imaging test to correctly rule out cancer. It is computed by dividing the number of patients who test negative on the imaging test by the number of patients who were actually negative via the gold standard test.

T staging — In the AJCC TNM system, T staging indicates the primary tumor size and/or spread. T0 means there is no tumor; TX means unknown size/spread; Tis means carcinoma in situ; T1 means the tumor size is < 2 cm, and the tumor is confined to the pancreas; T2 means the tumor size is > 2 cm, and the tumor is confined to the pancreas; T3 means the tumor has extended outside the pancreas but not to nearby arteries; and T4 means the tumor has extended outside the pancreas to nearby arteries.[20]

VII. Summary of Protocol Amendments

In the event of protocol amendments, the date of each amendment will be accompanied by a description of the change and the rationale.

VIII. Review of Key Questions

For all EPC reviews, Key Questions were reviewed and refined as needed by the EPC with input from Key Informants and the Technical Expert Panel (TEP) to ensure that the questions are specific and explicit about what information is being reviewed. In addition, the Key Questions were posted for public comment and finalized by the EPC after review of the comments.

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 health care 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 Task Order Officer (TOO) and the EPC work to balance, manage, or mitigate any potential conflicts of interest identified.

X. Technical Experts

Technical Experts comprise a multidisciplinary group of clinical, content, and methodological experts who provide input in defining populations, interventions, comparisons, or outcomes, as well as identifying particular studies or databases to search. They are selected to provide broad expertise and perspectives specific to the topic under development. Divergent and conflicted opinions are common and perceived as healthy scientific discourse that results in a thoughtful, relevant systematic review. Therefore, study questions, design, and/or 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 or contribute to the writing of the report and 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. Peer-review comments on the preliminary draft of the report are considered by the EPC in preparation of the final report. Peer Reviewers do not participate in writing or editing of the final report or other products. The synthesis of the scientific literature presented in the final report does not necessarily represent the views of individual reviewers. The dispositions of the peer-review comments are documented and will, for CERs and Technical Briefs, be published 3 months after the publication of the Evidence Report.

Potential 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. HHSA 290-2012-00011-I from the Agency for Healthcare Research and Quality (AHRQ), the U.S. Department of Health and Human Services. The TOO 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 AHRQ or the U.S. Department of Health and Human Services.

Appendix A: Example Search Strategy

Set Number	Concept	Search Statement
1	Pancreatic Cancer/pancreas	exp pancreatic neoplasms/ or pancreas cancer/ or pancreatic cyst/
2		((exp pancreas/ or exp pancreas, exocrine/) and (neoplasm/ or adenocarcinoma/ or neoplasms/)) or pancreas tumor/ or Pancreatic neuroendocrine tumor/
3		(Pancrea$ adj3 (mass$ or cancer$ or tumo?r$ or sarcoma$ or neoplasm$ or carcinoma$ or adenocarcinoma$)).mp.
4		neuroendocrine tumors/ and pancrea$.ti,ab.
5		Pancreatic neoplasms
6		4 not 5
7		1 or 2 or 3 or 6
8	Imaging	Endosonography/ or exp fluorodeoxyglucose F18/ or exp magnetic resonance imaging or exp positron-emission tomography/ or exp tomography, x-ray computed/ or exp ultrasonography/ or endoscopic echography/ or ultrasound/
9		Computer assisted emission tomography/ or computer assisted tomography/ or nuclear magnetic resonance imaging/ or positron emission tomography/
10		Exp Diagnostic imaging/ or exp magnetic resonance imaging/ or pancreas/ra or pancreas/us or pancreatic ducts/ra or pancreatic ducts/us
11		("computed tomography" or "positron emission" or "positron-emission" or "magnetic resonance" or "endoscopic ultraso$" or "computer assisted emission tomography" or "computer assisted tomography").ti,ab.
12		(SDCT or MDCT or FDG-PET or CT or PET or PET-CT or MRI or EUS).mp.
13		"Fludeoxyglucose positron emission tomography".mp.
14		((single?detector or multi?detector or multi?dimensional) adj (CT or "computed tomography")).mp.
15		(Endoscop$ adj (ultrasound or ultrasonograph$ or echograph$)).ti,ab.
16		Or/8-15
17	Pancreatic cancer and imaging	7 and 16
18	Diagnosis	exp pancreatic neoplasms/di or pancreatic cyst/di
19		Diagnos$.mp. or early diagnosis/ or diagnosis, differential/ or diagnosis/ or neuroendocrine tumor/di [Diagnosis] or

Set Number	Concept	Search Statement
20		pancreatic neoplasms/di, pa, us [Diagnosis, Pathology, Ultrasonography] or pancreas tumor/di [Diagnosis] or pancreas cancer/di
21		chronic pancreatitis/di [Diagnosis]
22		pancreatitis, chronic/di, pa, us [Diagnosis, Pathology, Ultrasonography]
23		Or/18-22
24	Pancreatic cancer and imaging and diagnosis	17 and 23
25	Staging	Cancer staging/ or Neoplasm staging/
26		((mass$ or cancer$ or tumo?r$ or sarcoma$ or neoplasm$ or carcinoma$ or adenocarcinoma$) and (biopsy or category or classification or classify or detect$ or stage or diagnos$ or staging)).ti,ab.
27		((mass$ or cancer$ or tumo?r$ or sarcoma$ or neoplasm$ or carcinoma$ or adenocarcinoma$) and ("fine needle" or fine-needle or FNA or FNB or specimen or sample or sampling)).ti,ab.
28		Or/25-27
29	Pancreatic cancer and imaging and staging	17 and 28
30	Diagnosis test – accuracy, specificity	17 and (exp diagnosis/ or di.fs.)
31		17 and (receiver operating characteristic/ or ROC curve/ or diagnostic accuracy/ or accuracy/)
32		17 and (precision or sensitivity or specificity or predict$ or forecast$ or likelihood or ((false or true) adj (positive or negative))).mp.
33		17 and (predictive value of tests/ or exp diagnostic errors/ or exp diagnostic error/)
34	Pancreatic cancer and imaging and diagnostic testing	Or/29-33
35	Screening	mass screening/ or early detection of cancer/ or population surveillance/

Set Number	Concept	Search Statement
36		Cancer screening/ or cancer risk/ or risk factors/
37		(risk$ or screen$ or hereditary or inherit$ or gene$ or family history).ti,ab.
38		or/ 35-37
39	Pancreatic cancer and imaging and screening	17 and 38
40	Prognosis/survival	incidence/ or mortality/ or follow up studies/ or prognos$.mp. or predict$.mp. or course$.mp. or (first and episode).ti,ab. or cohort.ti,ab.
41		Exp "prediction and forecasting"/ or exp prognosis/ or exp survival rate/ or surviv$.mp.
42		40 or 41
43	Pancreatic cancer and imaging and prognosis/survival	17 and 42

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

Modified QUADAS Instrument[33] for diagnostic test performance studies

1. Did the study enroll all, consecutive, or a random sample of patients?
2. Were the patient inclusion/exclusion criteria applied consistently to all patients?
3. Was the study affected by obvious spectrum bias?
4. Were there two or more readers?
5. If there were two or more readers, did the study account for inter-reader differences
6. Is the reference standard likely to correctly classify the target condition?
7. Were reader(s) of the diagnostic test of interest blinded to the results of the reference standard?
8. Were reader(s) of the reference standard blinded to the results of the diagnostic test of interest?
9. Were patients assessed by a reference standard regardless of the test's results?
10. Were all patients assessed by the same reference standard regardless of the test's results?
11. If the study reported data for a single diagnostic threshold, was the threshold chosen *a priori*?
12. Were the study results unaffected by intervening treatments or disease progression/regression?
13. Were at least 85% of the enrolled patients accounted for?

Additional risk-of-bias items for studies comparing two or more imaging tests

1. Was prior experience with the test (technicians, readers) similar for the two imaging tests being compared in the study?

2. Were the imaging tests performed within one month of each other (to avoid the possibility that the patient's true condition changed between tests)?3. 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)?

4. Did the interpreters have the same other information available at the time of interpretation for the imaging tests (other clinical information, 3rd test result)?

5. 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)?

6. 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)?

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 ≤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 ≥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 ≤15% difference in the length of followup for the two groups?
20. Did ≥85% of the patients complete the study?
21. Was there a ≤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?

Modified AMSTAR Instrument[32,43] for systematic reviews

1. Was an *a priori* design or protocol provided?
2. Was a comprehensive search strategy performed?

2a. Was this strategy appropriate to address the relevant Key Question of the CER?

3. Was a list of included and excluded studies provided?
4. Was the application of inclusion/exclusion criteria unbiased?

4a. Are the inclusion/exclusion criteria appropriate to address the relevant Key Question of the CER?

5. Was there duplicate study selection and data extraction?
6. Were the characteristics of the included studies provided?

7. Was the individual study quality assessed?
7a. Was the method of study quality assessment consistent with that recommended by the Methods Guide?
7b. Was the scientific quality of the individual studies used appropriately in formulating conclusions?
8. Were the methods used to combine the findings of studies appropriate?
9. Was the likelihood of publication bias assessed?
10. Have the authors disclosed conflicts of interest?