Evidence-based Practice Center Systematic Review Protocol

Project Title: Procalcitonin for Diagnosis and Management of Sepsis

Amendment Date(s) if applicable:

(Amendments Details–see Section VII)

I. Background and Objectives for the Systematic Review

Sepsis is a condition with high morbidity and mortality in which clinical diagnostic criteria lack sensitivity and specificity. Rapid diagnosis of sepsis and early initiation of antibiotic and goal-directed therapies have demonstrated a reduction in mortality. Conversely, overuse and misuse of antibiotics can result in adverse events and add to the increasing problem of antibiotic resistance. Several serum biomarkers have been identified in recent years with potential uses to help diagnose local and systemic infections; differentiate bacterial from viral or fungal infections; prognosticate; and guide antibiotic therapy. Among these potential uses of serum biomarkers, there is particular interest in finding a biomarker for diagnosis of sepsis. Currently, there are at least 178 serum biomarkers that have potential roles in the management of patients with infections, and 34 have been studied specifically for the diagnosis of sepsis. The serum biomarker that has been most extensively studied recently is procalcitonin, and the current literature suggests that it may prove to be the most useful biomarker for infections[1,2].

Procalcitonin is the prohormone precursor of calcitonin that is expressed primarily in C-cells of the thyroid gland and to a smaller extent in neuroendocrine tissue of other organs, such as lungs and intestines. The final step in conversion of procalcitonin to calcitonin is inhibited by various cytokines and bacterial endotoxin, and therefore, high levels of cytokines and/or bacterial endotoxin cause procalcitonin levels to rise. Cytokines are released non-specifically in response to inflammation and infection, but endotoxins are released specifically during bacterial infections since they are derived primarily from Gram negative bacterial cell walls. There is some evidence that procalcitonin is more specific for bacterial infections with serum levels rising and falling more rapidly in bacterial infection. Serum levels of procalcitonin were recognized to be elevated in patients with infections in the early 1990's, and since that time, numerous studies have investigated the potential roles of procalcitonin for diagnosis and management of local and systemic infections[3,4,5].

Procalcitonin's primary diagnostic utility is thought to be in establishing the presence of bacterial infections. Viruses, parasites, and fungi can increase procalcitonin levels due to systemic inflammation, but higher levels of procalcitonin have been demonstrated to specifically occur with bacterial infections, with the highest levels seen in bacterial sepsis. The diagnostic utility of procalcitonin is limited in fungal infections because the levels do not rise until 1-2 days after the onset of infection. A greater increase in

procalcitonin levels would be anticipated in Gram-negative vs. Gram-positive bacterial infections due to the release of endotoxin from Gram-negative bacterial cell walls; however, few studies have demonstrated higher levels of procalcitonin with Gram-negative compared to Gram-positive bacterial infections[5]. Procalcitonin appears to be a promising serum biomarker for infection, but its exact utility in diagnosing and managing patients with suspected infections remains unclear.

The FDA has approved at least three procalcitonin quantitative assays that are commercially available, but the optimal approach to laboratory testing of procalcitonin has yet to be clarified. Quantitative and qualitative (semi-quantitative) assays are currently available for measuring procalcitonin. The qualitative tests use test strips, are rapid (results available in <30 minutes), and are designed for point-of-care testing. The quantitative tests use luminescence immunoassay, are slower (results available in a few hours), and are designed for once or twice daily batch testing. Most studies supporting the use of procalcitonin have utilized the quantitative test which is neither rapid nor available at the bedside. Whether or not the semi-quantitative test will yield similar results to the quantitative test is unknown[4,9]. The analytic validity of quantitative and qualitative procalcitonin testing is important but is beyond the scope of this comparative effectiveness review and will not be addressed.

Although the utility of procalcitonin has been reviewed in meta-analyses, these reviews have been limited to select populations and most have analyzed the results of observational studies. Other reviews may not have used systematic methods or investigated diagnostic or therapy-monitoring uses. The most recent meta-analysis of the effects of procalcitonin-guided therapy in patients with infections included 7 randomized trials published through November of 2008[3]. Since that time, the number of trials studying procalcitonin-guided therapy has at least doubled, and there are at least 7 additional randomized trials that were not included in the prior review. Even though our understanding of procalcitonin is still evolving, clinicians have already begun to request laboratories perform procalcitonin measurements, and therefore, a systematic review of the major uses of procalcitonin is needed at the present time. Furthermore, a comprehensive review looking at all the potential uses of procalcitonin will identify the areas that require further prospective investigation and will serve as a roadmap for future research. The following Key Clinical Questions and Analytic Framework outline the approach and key issues to address in this review.

II. The Key Questions

During the topic posting period, four general comments were received. The comments varied greatly from being supportive to being skeptical of procalcitonin's utility in diagnosing and managing infections. The contrasting opinions about procalcitonin's utility underscore the importance of performing a formal comparative-effectiveness review. One comment questioned whether or not pediatric populations will be included in this review, and it should be noted that pediatric populations have not been excluded from this review. Studies have explored the use of procalcitonin in children with suspected infections, such as neonatal sepsis, urinary tract infections, and meningitis.

The utility of procalcitonin to screen for bacterial skin colonization or diagnose heat stroke are beyond the scope of this review and will not be included. No changes were made to the key clinical questions based on the public comments.

Key Question 1: In selected populations of patients with suspected local or systemic infection, what are the effects of using procalcitonin measurement plus traditional indicators of infection to guide initiation of antibiotic therapy, when compared to traditional indicators of infection alone, on:

- Intermediate outcomes such as diagnostic accuracy for occurrence of infection based on gold-standard test defined by individual study protocols, and timing of initiation of antibiotic therapy;

- Adverse events of testing such as pain, local bleeding or infection; and

- Health outcomes such as morbidity, mortality, function, quality of life and adverse events of therapy?

Key Question 2: In selected populations of patients who are being treated for local or systemic infection, what are the effects of procalcitonin measurement plus clinical signs/symptoms and other laboratory findings to guide decisions to continue or change therapy, when compared to decisions guided by clinical signs/symptoms and other laboratory findings alone, on:

- Intermediate outcomes such as changes in patient management, duration of antibiotic therapy, length of stay and response to therapy;

- Adverse events of testing such as persistent or recurrent infection, antibiotic resistance; and

- Health outcomes such as morbidity, mortality, function, quality of life and adverse events of therapy?

1. **Procalcitonin as a Diagnostic Indicator to Initiate Early Therapy in Patients with Suspected Infection**

    Population(s):

    - Adult patients with suspected infection including, but not limited to, the following:
        - Local infections
            - Acute exacerbation of chronic obstructive pulmonary disease
            - Pneumonia
            - Surgical site infection
            - Osteomyelitis
        - Systemic infections
            - Neutropenic fever

- Bacteremia
- Sepsis
- Septic shock

- Pediatric patients with suspected infection including, but not limited to, the following:
    - Local infections
        - Pneumonia
        - Urinary tract infection
        - Meningitis
    - Systemic infections
        - Neutropenic fever
        - Bacteremia
        - Sepsis
        - Septic shock

Interventions: Procalcitonin quantitative assay and semi-quantitative assay plus traditional indicators of infection (e.g., fever, leukocytosis, pyuria)

Comparators: Traditional indicators of infection (e.g., fever, leukocytosis, pyuria)

Outcomes:

- Intermediate outcomes: diagnostic accuracy for occurrence of infection, antibiotic exposure, duration of antibiotic therapy, length of stay
- Health outcomes: morbidity, mortality, function, and quality of life as measured by validated scales
- Adverse events: Pain, local bleeding, local infection, persistent or recurrent infection, antibiotic resistance

Timing:

- 3 months

Settings:

- Outpatient: Ambulatory clinics, urgent care centers
- Inpatient: Hospital wards, intensive care units, emergency departments

2. **Procalcitonin as an Indicator of Response to Therapy**
    Population(s):

- Adult and pediatric patients being treated for local infection including, but not limited to, the following:
  - Acute exacerbation of chronic obstructive pulmonary disease
  - Pneumonia
  - Osteomyelitis
- Patients being treated for systemic infection including, but not limited to, the following:
  - Sepsis

Interventions: Therapy (type, duration) guided by procalcitonin plus clinical signs/symptoms (e.g., fever, pain) and other laboratory findings (e.g., leukocytosis, C-reactive protein, bacteremia)

Comparators: Therapy (type, duration) guided by clinical signs/symptoms (e.g., fever, pain) and other laboratory findings (e.g., leukocytosis, C-reactive protein, bacteremia)

Outcomes:
- Intermediate outcomes: changes in patient management, duration of antibiotic therapy, length of stay, antibiotic exposure, accuracy for response to therapy
- Health outcomes: morbidity, mortality function, and quality of life as measured by validated scales
- Adverse events: Pain, local bleeding, local infection, persistent or recurrent infection, antibiotic resistance
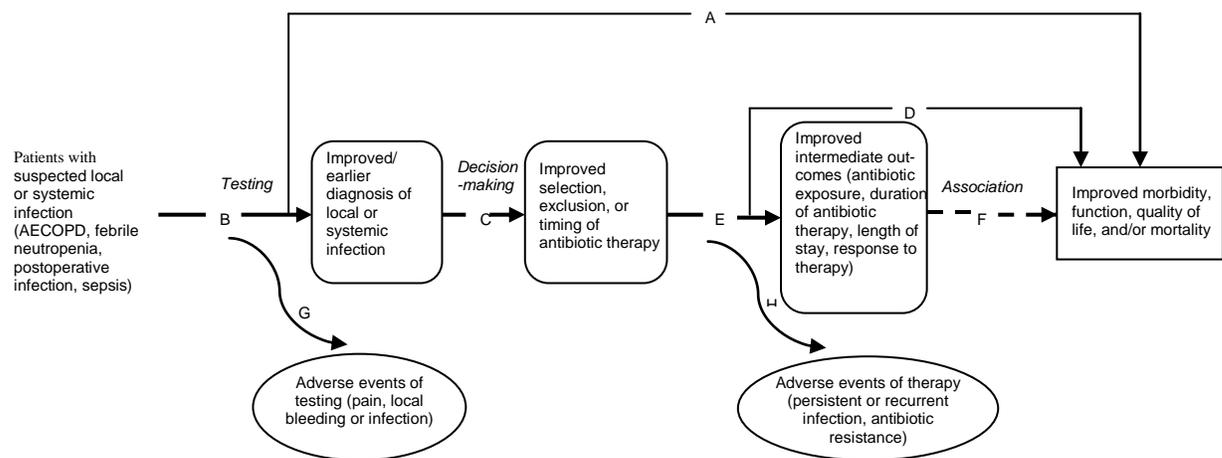
Timing:
- 3 months

Settings:
- Outpatient: Ambulatory clinics, urgent care centers
- Inpatient: Hospital wards, intensive care units, emergency departments

III. Analytic Framework

Figure 1

Procalcitonin as Diagnostic Indicator for Infection and as an Indicator of Response to Therapy

Abbreviations:

AECOPD = acute exacerbation of chronic obstructive pulmonary disease

Figure 1. This figure depicts the potential impact of use of procalcitonin on both intermediate outcomes and health outcomes. Direct evidence of the impact of testing on health outcomes is shown by link A (morbidity, function, quality of life and/or mortality) and link G (adverse events of testing). Indirect evidence would have to be assembled in the absence of controlled trials of the effects of testing on health outcomes. An early link in an indirect chain of evidence concerns the diagnostic accuracy of testing (B). Link C addresses whether test results influence decisions regarding therapy, which may have an impact on health outcomes (link D) or intermediate outcomes (link E). Intermediate outcomes, such as antibiotic exposure, duration of antibiotic therapy, length of stay and response to therapy, may have an association with health outcomes (link F). Link H focuses on the adverse events of therapy.

IV. Methods

Practices to be followed in this review will be derived from the Methods Guide for Comparative Effectiveness Reviews and its subsequent updates.

A. Criteria for Inclusion/Exclusion of Studies in the Review - We will include randomized, controlled studies and non-randomized, comparative studies (observational, case-control, and cohort studies) of populations, comparisons, interventions, and outcomes that were not adequately studied in controlled trials.   We will also use observational studies to assess comparative effectiveness in populations not well represented in randomized controlled trials. To classify observational study designs, we used the system developed by Briss and colleagues[10].

B. Searching for the Evidence: Literature Search Strategies for Identification of Relevant Studies to Answer the Key Questions - The following databases will be searched for citations. The search will be limited to English-language references because our EPC's experience in past projects that included non-English references did not yield high quality information that justified the resources required for translation. Furthermore, the search will be limited to literature published after 1990 which is approximately ten years before the use of modern serum biomarkers began to appear in the medical literature in the late 1990's.

- MEDLINE® (January 1, 1990, to present)

- EMBASE® (January 1, 1990, to present)

- Cochrane Controlled Trials Register (no date restriction)

To identify systematic reviews, we will search MEDLINE®, the Cochrane Database of Systematic Reviews, and the Web sites of the National Institute for Clinical Excellence, Guidelines.gov, and the NHA Health Technology Assessment Programme. We will use results from previously conducted meta-analyses and systematic reviews when appropriate.

Our search strategy will use the National Library of Medicine's Medical Subject Headings (MeSH) keyword nomenclature developed for MEDLINE® and adapted for use in other databases.   The searches will be limited to humans. We will search MEDLINE® for randomized controlled trials, non-randomized comparative studies,

and case series using the search term Procalcitonin (text word) AND "Systemic Inflammatory Response Syndrome"[Mesh] OR "Pulmonary Disease, Chronic Obstructive"[Mesh] OR "Surgical Wound Infection"[Mesh] OR "Critical Illness"[Mesh]) OR "Intensive Care"[Mesh] OR "Postoperative Complications"[Mesh] OR "Intensive Care Units"[Mesh] OR (("Neutropenia"[Mesh] OR neutropenia) AND febrile) OR sepsis OR septic OR "systematic inflammatory response syndrome" OR ICU OR "critically ill" OR "intensive care unit" OR "postoperative infection" AND Human AND English.

We will search EMBASE® for randomized controlled trials, non-randomized comparative studies, and case series using the search term Procalcitonin (text word) AND 'sepsis'/exp OR septic OR 'systemic inflammatory response syndrome'/exp OR 'copd'/exp OR 'chronic obstructive pulmonary disease'/exp OR 'febrile neutropenia'/exp OR 'postoperative infection'/exp OR 'postoperative infections'/exp OR 'postoperative complications'/exp OR 'post-surgical infection' OR 'post-surgical infections' OR 'critically ill'/exp OR icu OR 'intensive care'/exp OR 'intensive care units'/exp AND humans .

We will also search indexed, electronically searchable conference abstracts by subject heading for some, but not all, of the following conferences from the past 5 years: ICAAC (Interscience Conference on Antimicrobial Agents and Chemotherapy), Infectious Disease Society of America, American College of Physicians, American College of Chest Physicians, Pediatric Infectious Diseases Society, European Society of Clinical Microbiology and Infectious Diseases, International Society of Infectious Diseases, European Society of Infectious Diseases, British Society of Infectious Diseases, Australasian Society of Infectious Diseases, International Sepsis Forum, European Society of Intensive Care Medicine, American Association for Clinical Chemistry, College of American Pathology.

The Technical Expert Panel and individuals and organizations providing peer review will be asked to inform the project team of any studies relevant to the key questions that were not included in the draft list of selected studies.

We will review Scientific Information Packets from the Scientific Resource Center and gray literature from conferences, the Food and Drug Administration website and clinicaltrials.gov.

Search results will be stored in an EndNote9® or ProCite® database. Using the study selection criteria for screening titles and abstracts, a single reviewer marked each citation as either: (1) eligible for review as full-text articles; (2) ineligible for full-text review; or (3) uncertain. Citations marked as uncertain will be reviewed by a second reviewer and resolved by consensus opinion, with a third reviewer to be consulted if necessary. Using the final study selection criteria, review of full-text articles will be conducted in the same fashion to determine inclusion in the systematic review. Records of the reason for exclusion for each paper retrieved in full-text, but excluded from the review, will be kept in the EndNote9® or ProCite® database.

C. Data Abstraction and Data Management

Data Elements- The data elements following will be abstracted, or recorded as not reported, from intervention studies. Data elements to be abstracted will be defined in consultation with the TEP. They will include the following:

- Quality Assessment:
  - Number of participants and flow of participants through steps of study
  - Treatment allocation methods (including concealment)
  - Use of blinding
  - Prospective vs. retrospective
  - Use of independent outcome assessor
  - Additional elements are described below under Assessment of Methodological Quality of Individual Studies
- Assessment of Applicability & Clinical Diversity:
  - Patient characteristics, including
    - Age
    - Sex
    - Race/ethnicity
    - Disease and type
    - Disease duration
    - Other prognostic characteristics (e.g., comorbidities and other potential confounders and/or effect modifiers)
    - Setting
      - Outpatient
      - Inpatient
  - Diagnostic and Treatment characteristics, including
    - Procalcitonin assay type
    - Other measured indicators of sepsis or of response to treatment (e.g., fever, white blood cell count)
    - Decision-making for diagnosis and/or treatment (e.g. when to administer antibiotic therapy)
    - Antibiotic usage during study period
    - Duration of observation
    - Other treatment modalities
- Outcome Assessment:
  - Identified primary outcome
  - Identified secondary outcomes
  - Response criteria
  - Follow-up frequency and duration
  - Data analysis details:

- Statistical analyses (statistical test/estimation results)
  - Test used
  - Summary measures
  - Sample variability measures
  - Precision of estimate
  - p values
- Regression modeling techniques
  - Model type
  - Candidate predictors and methods for identifying candidates
  - Univariate analysis results
  - Selected predictors and methods for selecting predictors
  - Testing of assumptions
  - Inclusion of interaction terms
  - Multivariable model results
  - Discrimination or validation methods and results
  - Calibration or "goodness-of-fit" results
- The same abstraction tables will be used for comparative and single-arm studies, although some elements may not apply to the latter (e.g., description of control group).

Evidence Tables-Templates for evidence tables will be created in Microsoft Access®. One reviewer will perform primary data abstraction of all data elements into the evidence tables, and a second reviewer will review articles and evidence tables for accuracy. Disagreements will be resolved by discussion, and if necessary, by consultation with a third reviewer. When small differences occur in quantitative estimates of data from published figures, the values will be obtained by the two reviewer average.

D. Assessment of Methodological Quality of Individual Studies
- **Definition of Ratings Based on Criteria-**In adherence with EPC Methods Guide developed by the AHRQ EPC Program[11], the general approach to grading individual comparative studies will be performed using a method used by the U.S. Preventive Services Task Force (USPSTF)[12]. The quality of the abstracted studies and the body of evidence will be assessed by two independent reviewers. Discordant quality assessments will be resolved with input from a third reviewer, if necessary.
- The quality of studies will be assessed on the basis of the following criteria:
  - Initial assembly of comparable groups: adequate randomization, including concealment and whether potential confounders (e.g., other concomitant care) were distributed equally among groups
  - Maintenance of comparable groups (includes attrition, crossovers, adherence, contamination)
  - Important differential loss to follow-up or overall high loss to follow-up

- o Measurements: equal, reliable, and valid (includes masking of outcome assessment)
  - o Clear definition of interventions
  - o All important outcomes considered
  - o Analysis: adjustment for potential confounders, intention-to-treat analysis
- The rating of intervention studies encompasses the three quality categories described here.
  - o *Good:* Meets all criteria; comparable groups are assembled initially and maintained throughout the study (follow-up at least 80 percent); reliable and valid measurement instruments are used and applied equally to the groups; interventions are spelled out clearly; all important outcomes are considered; and appropriate attention is given to confounders in analysis. In addition, for randomized, controlled trials, intention to treat analysis is used.
  - o *Fair:* Studies graded "fair" if any or all of the following problems occur, without the fatal flaws noted in the "poor" category below: In general, comparable groups are assembled initially but some question remains whether some (although not major) differences occurred with follow-up; measurement instruments are acceptable (although not the best) and generally applied equally; some but not all important outcomes are considered; and some but not all potential confounders are accounted for. Intention-to-treat analysis is done for randomized, controlled trials.
  - o *Poor:* Studies graded "poor" if any of the following fatal flaws exists: Groups assembled initially are not close to being comparable or maintained throughout the study; unreliable or invalid measurement instruments are used or not applied at all equally among groups (including not masking outcome assessment); and key confounders are given little or no attention. For randomized, controlled trials, intention-to-treat analysis is lacking.
- The quality of included nonrandomized comparative intervention studies will be also assessed based on a selection of items proposed by Deeks et al.[13] to inform the USPSTF approach, as follows:
  - o Was sample definition and selection prospective or retrospective?
  - o Were inclusion/exclusion criteria clearly described?
  - o Were participants selected to be representative?
  - o Was there an attempt to balance groups by design?
  - o Were baseline prognostic characteristics clearly described and groups shown to be comparable?
  - o Were interventions clearly specified?
  - o Were participants in treatment groups recruited in the same time period?
  - o Was there an attempt by investigators to allocate participants to treatment groups in an attempt to minimize bias?
  - o Were concurrent/concomitant treatments clearly specified and given equally to treatment groups?

- o Were outcome measures clearly valid, reliable and equally applied to treatment groups?
  - o Were outcome assessors blinded?
  - o Was the length of follow-up adequate?
  - o Was attrition below an overall high level (less than 20 percent)?
  - o Was the difference in attrition between treatment groups below a high level (less than 15 percent)?
  - o Did the analysis of outcome data incorporate a method for handling confounders such as statistical adjustment?
- The quality of included diagnostic accuracy studies will be assessed by the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool, which underwent a rigorous development process by Whiting, Rutjes, Dinnes, et al.[14] and includes the following items:
  - o Was the spectrum of patients representative of the patients who will receive the test in practice?
  - o Were selection criteria clearly described?
  - o Is the reference standard likely to classify the target condition correctly?
  - o Is the period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?
  - o Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis?
  - o Did patients receive the same reference standard regardless of the index test result?
  - o Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)?
  - o Was the execution of the index test described in sufficient detail to permit replication of the test?
  - o Was the execution of the reference standard described in sufficient detail to permit replication of the reference standard?
  - o Were the index test results interpreted without knowledge of the results of the reference standard?
  - o Were the reference standard results interpreted without knowledge of the results of the index test?
  - o Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?
  - o Were uninterpretable/intermediate test results reported?
  - o Were withdrawals from the study explained?
- The quality of included single-arm intervention studies will be assessed based on a set of study characteristics proposed by Carey and Boden[15], as follows:
  - o Clearly defined question
  - o Well-described study population
  - o Well-described intervention
  - o Use of validated outcome measures
  - o Appropriate statistical analyses

- o Well-described results
- o Discussion and conclusion supported by data
- o Funding source acknowledged
- The quality of included predictive studies will be assessed based on an approach applied to a systematic review of HER2 testing for breast cancer and other solid tumors (Table 1) [16].

Table 1. Hierarchy of study design and conduct for assessing procalcitonin measurement for prediction of outcome

More informative	Randomized trial, randomization stratified by procalcitonin level OR patients randomized to procalcitonin-guided treatment or non-procalcitonin-guided treatment
↑	Randomized trial, prespecified multivariate subgroup analysis
↑	Randomized trial, post-hoc multivariate subgroup analysis
Continuum	Randomized trial, treatment by procalcitonin level subgroup analysis
↓	Single-arm study, prespecified multivariate analysis
↓	Single-arm study, post-hoc multivariate analysis
Less informative	Single-arm study, univariate analysis

- Table 1 shows the framework for evaluating how informative different designs and analytic strategies would be to predictions of outcomes according to different categories or levels of predictive factors. The most informative scenario would be a trial in which randomized assignment to treatment groups would be stratified by predictive factor level or patients were randomized to receive treatment guided by predictive factor or not. An adequately powered stratified randomization would allow valid inferences of treatment by predictive factor interactions. Randomized trials generally are preferred because they convey the possibility of determining differences in the relative efficacy of two treatments, whereas single-arm studies can only assess the association between predictive factor and outcomes after a single treatment regimen. Subgroup analyses in randomized trials should ideally assess the significance of treatment effect interactions. Prespecified subgroup analyses guard against the problems of data dredging.

E. **Data Synthesis-** Whether or not this evidence review will incorporate formal data synthesis using meta-analysis will be determined after completing the formal literature search. If meta-analysis can be performed, subgroup and sensitivity analyses will be based on assessment of clinical diversity in available studies. The EPC Methods Guide developed by the AHRQ EPC Program[18] will be used to rate the strength of the overall body of evidence

F. **Grading the Evidence for Each Key Question**-Applicability of findings in this review will be assessed within the EPICOT framework (Evidence, Population, Intervention, Comparison, Outcome, Timestamp)[17]. Selected studies will be assessed for relevance against target populations, interventions of interest, and

outcomes of interest. The system used for rating the strength of the overall body of evidence was developed by the AHRQ EPC Program for the EPC Methods Guide, based on a system developed by the GRADE Working Group[18]. This system explicitly addresses the following domains: risk of bias, consistency, directness, and precision. Grade of evidence strength is classified into the following four categories:

- High-High confidence that the evidence reflects the true effect. Further research is very unlikely to change our confidence in the estimate of effect.
- Moderate-Moderate confidence that the evidence reflects the true effect. Further research may change our confidence in the estimate of effect and may change the estimate.
- Low-Low confidence that the evidence reflects the true effect. Further research is likely to change our confidence in the estimate of effect and is likely to change the estimate.
- Insufficient-Evidence is either unavailable or does not permit estimation of an effect.
- Additional domains includingstrength of association, publication bias, coherence, dose-response relationship, and residual confounding will be addressed if appropriate

Specific outcomes and comparisons to be rated will depend on the evidence found in the literature review. The grade rating will be made by independent reviewers and disagreements will be resolved by consensus adjudication.

V. References

1. Pierrakos C, Vincent JL. Sepsis biomarkers: a review. Critical Care 2010;14:R15.

2. Marshall JC, Reinhart K. Biomarkers of sepsis. Critical Care Medicine 2009;37(7):2290-8.

3. Tang H, Huang T, Jing J, et al. Effect of procalcitonin-guided treatment in patients with infections: a systematic review and meta-analysis. Infection 2009;37:497-507.

4. Jones AE, Flecht JF, Brown MD, et al. Procalcitonin test in the diagnosis of bacteremia: a meta-analysis. Annals of Emergency Medicine 2007;50(1):34-40.

5. Becker KL, Snider R, Nylen ES. Procalcitonin in sepsis and systemic inflammation: a harmful biomarker and a therapeutic target. British Journal of Pharmacology 2010;159:253-64.

6. Tang BM, Eslick GD, Craig JC, McLean AS. Accuracy of procalcitonin for sepsis diagnosis in critically ill patients: a systematic review and meta-analysis. Lancet Infectious Disease 2007;7:210-17.

7. Stolz A, Tamm M. Discriminate use of antibiotics for exacerbation of COPD. Current Opinion in Pulmonary Medicine. 2009;15:126-132.

8. Daniels JMA, Schoorl M, Snijders, Knol DL, Lutter R, Jansen HM, Boersma WG. Procalcitonin versus C-reactive protein as predictive markers of response to antibiotic therapy in acute exacerbations of COPD. Chest 2010.

9. Schneider HG, Lam QT. Procalcitonin for the clinical laboratory: a review. Pathology 2007;39(4):383-90.

10. Briss PA, Zaza S, Pappaioanou M et al. Developing an evidence-based Guide to Community Preventive Services--methods. The Task Force on Community Preventive Services. Am J Prev Med 2000;18(1 Suppl):35-43.

11. Agency for Healthcare Research and Quality. Methods Guide for Effectiveness and Comparative Effectiveness Reviews, Version 1.0 [Draft posted Oct. 2007]. Rockville, MD. Available at: http://effectivehealthcare.ahrq.gov/repFiles/2007_10DraftMethodsGuide.pdf. Accessed December 2009.

12. Harris RP, Helfand M, Woolf SH, et al. for the Third US Preventive Services Task Force. Current methods of the US Preventive Services Task Force: a review of the process. Am J Prev Med 2001;20(3 Suppl):21-35.

13. Deeks JJ, Dinnes J, D'Amico R, et al. Evaluating non-randomised intervention studies. Health Technol Assess 2003;7(27):iii-x, 1-173.

14. Whiting P, Rutjes AWS, Dinnes J, et al. Development and validation of methods for assessing the quality of diagnostic accuracy studies. Health Technol Assess 2004; 8(25):1-234.

15. Carey TS, Boden SD. A critical guide to case series reports. Spine 2003;28(15):1631-1634.

16. Seidenfeld J, Samson DJ, Rothenberg BM, Bonnell CJ, Ziegler KM, Aronson N. HER2 Testing to Manage Patients With Breast Cancer or Other Solid Tumors. Evidence Report/Technology Assessment No. 172. (Prepared by Blue Cross and Blue Shield Association Technology Evaluation Center Evidence-based Practice Center, under Contract No. 290-02-0026.) AHRQ Publication No. 09-E001. Rockville, MD: Agency for Healthcare Research and Quality. November 2008.

17. Brown P, Brunnhuber K, Chalkidou K, et al. How to formulate research recommendations. BMJ 2006;333(7572):804-806.

18. Owens DK, Lohr KN, Atkins D et al. AHRQ series paper 5: Grading the strength of a body of evidence when comparing medical interventions-- Agency for Healthcare Research and Quality and the Effective Health-Care Program. J Clin Epidemiol 2010;63(5):513-23.

VI. Definition of Terms

None

VII. Summary of Protocol Amendments

Section	Original Protocol	Revised Protocol	Rationale
II. Key Questions, Question 1	Intermediate outcomes, such as _diagnostic accuracy for occurrence of infection based on a gold-standard test defined by individual study protocols,_ and timing of initiation of antibiotic therapy?	Intermediate outcomes, such as _detection of infection_ and timing of initiation of antibiotic therapy?	This change is based on the fact that availability of direct evidence precludes the need for indirect evidence. Because randomized controlled trials were available that directly addressed the Key Questions, there was no need to review diagnostic accuracy studies. Diagnostic accuracy studies need to be reviewed in the absence of randomized trials addressing procalcitonin-guided therapy. Such studies would inform one link in an analytic framework, providing indirect evidence. We are encouraged that investigators chose to conduct randomized trials and support collection of direct evidence. In the section of the report focusing on research gaps, we will emphasize unresolved questions that could be addressed with direct evidence. Research gaps centered on indirect evidence will not be emphasized..
II. Key Questions, Question 1	Adverse events of _testing, such as pain, local bleeding, or infection_?	Adverse events _of testing and therapy (e.g., persistent or recurrent infection, and antibiotic resistance)?_	The change in wording related to adverse events reflects the importance of possible harms resulting from therapy chosen as a result of testing (i.e., inappropriately guided therapy). The harms of the testing procedure itself would relate to venipuncture.
II. Key Questions, Question 2	Adverse events of _testing, such as persistent or recurrent infection, and antibiotic_	Adverse events _of testing and therapy (e.g., persistent or recurrent infection, and antibiotic resistance)?_	

	resistance?		
II. Key Questions, PICOTS Framework, 1c. Comparators	*Traditional indicators of infection (e.g., fever, leukocytosis, pyuria)*	*Initiation of antibiotics based on traditional indicators of infection (e.g., fever, leukocytosis, pyuria), as described in guidelines*	The comparator is specified as initiation of antibiotics based on traditional indicators of infection as described in guidelines to clarify the true comparator for procalcitonin-guided initiation of antibiotics.
II. Key Questions, PICOTS Framework, 2c. Comparators	Therapy (type, duration) *guided by* clinical signs/symptoms (e.g., fever, pain), and other laboratory findings (e.g., leukocytosis, C-reactive protein, bacteremia).	Therapy (type, duration) *based on* clinical signs/symptoms (e.g., fever, pain), and other laboratory findings (e.g., leukocytosis, C-reactive protein, bacteremia), *as described in guidelines.*	The comparator is qualified as being therapy described in guidelines, and the statement is kept parallel with the previously mentioned comparator for Question 1.
IV. Methods, D. Assessment of Methodological Quality of Individual Studies; Definition of Ratings Based on Criteria	**Poor.** Studies are graded "poor" if any of the following fatal flaws exists: Groups assembled initially are not close to being comparable or maintained throughout the study; unreliable or invalid measurement instruments are used or not applied at all equally among groups; and key confounders are given little or no attention. For RCTs, intention-to-treat analysis is lacking.	**Poor.** Studies are graded "poor" if any of the following fatal flaws exists: Groups assembled initially are not close to being comparable or maintained throughout the study; unreliable or invalid measurement instruments are used or not applied at all equally among groups; and key confounders are given little or no attention. *Generally, lack of masked outcome assessment is considered a fatal flaw, but due to the nature of the interventions and comparators in this systematic review, it is not considered a fatal flaw.* Specifically, outcomes such as antibiotic use/duration and	Among the key outcomes considered in this review were antibiotic use/duration and hospital/ICU length of stay. These outcomes were determined by judgments made by treating clinicians who had knowledge of whether treatment was guided by procalcitonin or not. Detection bias, and therefore use of masked outcome assessors, would not be relevant to such outcomes. Nor would masked outcome assessors be relevant to mortality outcomes.

		hospital/ICU length of stay are determined by judgments made by treating clinicians who had knowledge of whether treatment was guided by procalcitonin or not. Detection bias, and therefore use of masked outcome assessors, would not be relevant to such outcomes. Nor would masked outcome assessors be relevant to mortality outcomes. For RCTs, intention-to-treat analysis is lacking.	
IV. Methods, D. Data Synthesis	Whether or not this evidence review will incorporate formal data synthesis, we will decide to perform a meta-analysis will be determined after completing the formal literature search. If a meta-analysis can be performed, subgroup and sensitivity analyses will be based on assessment of clinical diversity in available studies. The strength of the overall body of evidence will be rated by using the system described in the Methods	Whether or not this evidence review will incorporate formal data synthesis, we will decide to perform a meta-analysis will be determined after completing the formal literature search. The decision to pool will be based on whether there is a sufficient number of studies designed to ask similar questions and reporting simililarly defined outcomes. If a meta-analysis can be performed, subgroup and sensitivity analyses will be based on assessment of clinical diversity in available studies. The pooling method will involve inverse variance weighting and a random	The rationales for deciding to perform meta-analyis and the pooling technique were based on recommendations of the AHRQ Methods Guide for Comparative Effectiveness Reviews (accessible at: http://www.effectivehealthcare.ahrq.gov/ehc/products /243/554/MethodsGuide--ConductingQuantitativeSynthesis.pdf

	Guide.[10]	effects model. The strength of the overall body of evidence will be rated by using the system described in the Methods Guide.[10]	

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 assure that the questions are specific and explicit about what information is being reviewed. In addition, for Comparative Effectiveness reviews, 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 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 comprise a multi-disciplinary 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 health 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 nor 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 draft of the 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 three 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.