Comparative Effectiveness Research Review Disposition of Comments Report

Research Review Title: *Procalcitonin-Guided Antibiotic Therapy*

Draft review available for public comment from September 9, 2011 to October 7, 2011

Research Review Citation: Soni NJ, Samson DJ, Galaydick JL, Vats V, Pitrak DL, Aronson N. Procalcitonin-Guided Antibiotic Therapy. Comparative Effectiveness Review No. 78 (Prepared by the Blue Cross and Blue Shield Association Technology Evaluation Center Evidence-based Practice Center under Contract No. 290-2007-10058-I.) AHRQ Publication No. 12(13)-EHC124-EF. Rockville, MD: Agency for Healthcare Research and Quality. October 2012. Available at: www.effectivehealthcare.ahrq.gov/reports/final.cfm.

Comments to Research Review

The Effective Health Care (EHC) Program encourages the public to participate in the development of its research projects. Each comparative effectiveness research review is posted to the EHC Program Web site in draft form for public comment for a 4-week period. Comments can be submitted via the EHC Program Web site, mail or email. At the conclusion of the public comment period, authors use the commentators' submissions and comments to revise the draft comparative effectiveness research review.

Comments on draft reviews and the authors' responses to the comments are posted for public viewing on the EHC Program Web site approximately 3 months after the final research review is published. Comments are not edited for spelling, grammar, or other content errors. Each comment is listed with the name and affiliation of the commentator, if this information is provided. Commentators are not required to provide their names or affiliations in order to submit suggestions or comments.

The tables below include the responses by the authors of the review to each comment that was submitted for this draft review. The responses to comments in this disposition report are those of the authors, who are responsible for its contents, and do not necessarily represent the views of the Agency for Healthcare Research and Quality.

Source: http://effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=1273
Published Online: October 5, 2012

Commentator & Affiliation	Section	Comment	Response
PR (#7)	Executive Summary	The Executive Summary is clear and well-written. It provides a very good overview of key elements of the CER.	No response necessary
PR (#7)	Executive Summary	Page ES-3, Lines 23 - 26: There is reference to a 'comprehensive review' evaluating different patient populations ..... will serve as a roadmap for future research. What is the meaning of this sentence? It doesn't seem that it could be referring to this CER. So is there another effort actually being worked on or are the authors of this CER simply commenting on what the benefits of such a comprehensive review on PCT would be. If this statement is in the ES to rationalize the narrowing of focus to the CER topic and key questions, this sentence and some additional information could be more clearly stated as to what the authors intent was in including it.	The sentence is referring to how this CER will serve as a roadmap for determining future research needs, but because it is misleading, it has been revised as follows Further, the present review will identify the areas that require further prospective investigation and will provide suggested directions for future research
PR (#7)	Executive Summary	Table ES-1, Page ES-7: There could be different ways to read and interpret the options for the Conflict of Interest row. I read it that Yes means a COI exists, No means it does not, NR means that it was Not Recorded (and/or Not asked). Whether that's the right interpretation or I'm incorrect, it would be appropriate to describe in the table legend what is meant in each case.	We have amended the legend of the row to clarify this further: "Yes implies published paper reported existence of conflict of interest No implies published paper reported no existence of conflict of interest NR implies published paper did not report whether conflict of interest existed or not"
PR (#7)	Executive Summary	Table ES-2, Page ES-9: While I knew what the designations mean, I thought it would be helpful to refer to Table ES-2 in the ES (page ES-6, line 37) where the ES describes the elements of GRADE (B, C, D and P).	We disagree with the reviewer because it will not be appropriate to cross reference a result table in the methods chapter.
PR (#2)	Introduction	Introduction: The relevant issues are well described in the introduction. One issue that only comes up later in the discussion is the degree to which the use of PCT testing is merely a substitute for any of a number of available clinical algorithms for guiding treatment decisions. Few if any of the PCT trials compared the PCT algorithm to established treatment algorithms (for example, for acute respiratory infections). However, trials of these clinical algorithms, which largely utilize history and physical examination findings, often find effect sizes comparable to those reported for PCT. For example, studies of clinical algorithms to reduce antibiotic prescribing for acute respiratory infections often report a reduction in overall antibiotic use of about 20%, consistent with the PCT findings in this group.	Both PCT and antibiotic algorithms are only guides to clinical decision making which are individualized to specific patient. Addition of PCT to algorithm adds a quantifiable and measurable parameter to an algorithm. A few studies referenced established treatment guidelines, but majority were based on standard treatment per physician. We may speculate that PCT testing is equivalent to other algorithms for guiding treatment decisions but testing this idea requires finding studies that make comparisons of procalcitonin-guided therapy and therapy guided by other means. We have searched for and reviewed such studies and summarize our findings.

Commentator & Affiliation	Section	Comment	Response
PR (#6)	Introduction	Page 2, line 27 - 35. Agree with the statements about the importance of the analytic approach to PCT testing that is, yet, beyond the scope of this report. The subsequent discussion and rationalization for how the PCT results were viewed and interpreted is reasonable and effective. As described in the report and detailed in the abstraction tables, 16 of the 18 RCT's were truly using the same methodology for PCT testing with the 17th being a semi-quantitative version of this same PCT methodology and the 18th RCT not identifying how PCT was measured.	No response necessary
PR (#6)	Introduction	Again, PICOTS and key questions were clear and appropriate. Breaking down the included studies across five different patient populations was a good strategy. It allowed effective assessment within more narrowly defined populations as well as the ability to increase the strength of the evidence within a study population.	No response necessary
PR (#7)	Introduction	Definitely a needed contribution to the literature.	No response necessary
PR (#2)	Methods	The methods are very nicely presented and easy to follow. The separation of proximal, process measures and distal clinical outcomes is clearly described and appropriately summarized. The limited use of meta-analytic techniques was appropriate.	No response necessary
PR (#3)	Methods	Most screening abstracts were only done by one reviewer.	This is correct. We have also done extensive review of bibliographies of other reviews and included studies and feel confident that no comparative studies were missed.
PR (#3)	Methods	One of the exclusion criteria “Did not report primary data” needs more clarification.	This means that a paper did not report data on original research. This has been clarified.
PR (#3)	Methods	Definition of “Intention to treat analysis” – One important aspect of “Intention to treat analysis” is to analyze randomised controlled trials that compares patients in the groups to which they were originally randomly assigned.	This concept is part of our definition of intention-to-treat analysis. We will make this explicit.

Commentator & Affiliation	Section	Comment	Response
PR (#3)	Methods	Data synthesis: the paragraph described the data (quantitative) synthesis inadequately. It is written as if the review has NOT been done. More details are needed for the synthesis methods with rational whether or not to combine studies.	More detail has been added to the section on Data Synthesis. It now reads: We anticipated that the decision to incorporate formal data synthesis into this evidence review would be made after completing the formal literature search. Similarly we also anticipated that the decision to pool studies would be based if there were sufficient number of studies available that were designed to ask similar questions and reported similarly defined outcomes. Specifically, meta-analysis was performed if a minimum of three studies selected a similar population and reported an outcome that was measured similarly and with sufficient detail. An example of similarly measured outcomes included combining in-hospital mortality, 28 day mortality or 6 week mortality. Antibiotic use outcomes that were considered distinct included duration of antibiotic therapy in days, rate of antibiotic prescription as a proportion, and total antibiotic exposure in number of days on antibiotics per total patient-days for an entire group as a rate. Sufficient detail for an outcome measured as a mean, such as length of stay, requires reporting of standard deviations. When meta-analysis could be performed, subgroup and sensitivity analyses would be based on assessment of study-level clinical diversity in a sufficient number of available studies. Degree of statistical heterogeneity is reported by the I^2 statistic. The pooling method involves inverse variance weighting and a random effects model. Studies reporting zero events in one or both arm are excluded from pooling.
PR (#6)	Methods	Topic refinement was appropriately focused around the key questions and PICOTS to generate an analytic framework for PCT as a diagnostic indicator for infection and an indicator of response to therapy that made sense. Inclusion and exclusion criteria were justifiable. I had no other criteria to suggest. In general, search strategies were explicitly stated and logical although I have a comment on the PRISMA table for the Gray Literature search in the Results section below.	No response necessary
PR (#6)	Mehtods	Data extraction and management were effective as well as the detailed information, particularly in the abstraction tables, for the findings of each included study as well as the basis for rating the quality of each included study.	No response necessary

Commentator & Affiliation	Section	Comment	Response
PR (#6)	Methods	Page 13, Line 5: In describing how studies received a "poor" rating, they used the phrase fatal flaw. I understand that term although not totally clear on its use here. Is this just a convention that I'm unaware of or "technial jargon?" Fatal meaning that this is how a study ends up in the poor category? At any rate, the use of the term fatal flaw in this context seemed a little unusual.	Yes, in the USPSTF framework a fatal flaw is an error that automatically downgrades the quality of the study to "poor." In the methods chapter, examples of fatal flaws are listed, but in this report, the most common fatal flaw was a lack of intention-to-treat analyses.
PR (#6)	Methods	Overall, the descriptions of how the evidence was graded for each key question as well as the descriptions of the ratings definitions for individual study quality assessment was clear and understandable.	No response necessary
PR (#7)	Methods	This study should examine the grading of the Surviving Sepsis Campaign in 2008. This will be a continuing question no matter what the methodolgy used in this study. Using these criteria, the grading of evidence should be recommended.	Examining the grading criteria used in the Surviving Sepsis campaign is beyond the scope of this review. The criteria and methods used to assess the strength of evidence are based on the Effective Health Care Program Methods Guide chapter entitled: Grading the Strength of a Body of Evidence when Comparing Medical Interventions Link: http://www.effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and reports/?pageaction=displayProduct&productID=328
PR (#2)	Results	Overall, the results are well presented and contain the appropriate level of detail. In focusing on the 2 settings with the "positive" findings, I wonder whether the results could be further subdivided based on the heterogeneity of the target populations. Specifically, for the use of PCT to reduce antibiotic use in ICU populations, is the effect stable across diverse ICU populations, including those with and without mechanical ventilation? For the use of PCT to reduce initiation or continuation of antibiotics for acute respiratory tract infections, is the result consistent across inpatient and outpatient settings?	Past systematic reviews grouped patients that did not distinguish clinical populations, and we attempted to group patients into groups based on clinical condition, and also differentiate adult from pediatric populations. We pooled patients into clinically relevant groups based on the similarities of the groups. Subdividing the groups further would results in several sparsely populated groups. We would have liked to compare effects between those with versus those without mechanical ventilation, but data for such comparisons was lacking among available studies. RTI studies included only one within an inpatient setting, all others were either outpatient or in emergency departments so it is not possible to divide RTI more finely
PR (#3)	Results	Also in the results section, quantitative synthesis was not adequately done and there is not much assessment of statistical heterogeneity. In many places, the results across multiple studies were consistent enough and a combined estimate would provide more precise and conclusive results. I would recommend the investigators to assess the similarity among studies for each outcome and each population to see whether quantitative synthesis would help to summarize the evidence and provide more precise estimates.	As noted above, meta-analysis was performed if a minimum of three studies selected a similar population and reported an outcome that was measured similarly and with sufficient detail. The Results chapter sections describing meta-analyses consistently note the degree of statistical heterogeneity in each pooling.

Commentator & Affiliation	Section	Comment	Response
PR (#3)	Results	Critically Ill/Ventilator-Associated Pneumonia Antibiotic use -- the results seemed to be consistent. No quantitative synthesis was done. If the authors have a reason not doing this, they should provide a rational.	To give a clearer assessment of consistency and precision, we have added add a meta-analysis for antibiotic treatment duration.
PR (#3)	Results	Critically Ill/Ventilator-Associated Pneumonia In table 4 and other many tables in the report, the 95% CI for mean difference could be calculated from P value (Stolz 2009). Other information reported in the study may also be helpful to calculate the 95% or standard error.	We were unable to calculate 95% CI in this case as the authors reported only the median values along with inter quartile range. Mean value with standard deviation or standard error of the mean were not reported. Further, we have added text in the footnote of table 4 as well in other tables throughout the report to clarify if the reported values are mean or median.
PR (#3)	Results	Critically Ill/Ventilator-Associated Pneumonia Mortality -- Table 5, explain why two rows of data in Nobre, 2008 for 28-day mortality.	Row 1 is 28-day mortality Row 2 is in-hospital mortality This is clearly laid out in table 5.
PR (#3)	Results	Critically Ill/Ventilator-Associated Pneumonia Mortalty -- results of meta-analysis for short mortality (28-day or inhospital).	This comment is unclear. As noted above, we combined results of studies that reported similarly measured outcomes. We considered in-hospital mortality and 28 day mortality to be similarly measured, thereby allowing for pooling.
PR (#3)	Results	Critically Ill/Ventilator-Associated Pneumonia For figure 4, five studies were included. However, based on Table 5, there are only three studies reported results on 28-day or in hospital mortality. Hochreiter and Schroeder reported just mortality.	In the final report, we clarified in the meta-analysis (Figure 5) that we are pooling 28-day and in-hospital mortality from the critically ill/ventilator-associated pneumonia patients. Data from all 5 studies was pooled. Nobre, Stolz, and Bouadma reported 28-day mortality, and Nobre, Stolz, Hochreiter, and Schroeder reported in-hospital mortality. Note: the comment refers to Figure 4 which is Figure 5 in the final report.
PR (#3)	Results	Critically Ill/Ventilator-Associated Pneumonia Morbidity -- ICU LOS, again the 95% CI for Nobre could be calculated (and Stolz, too in table 6).	95% CI for Nobre, 2008 already exists in table 5. 95% CI for Hochreiter 2009 and Stolz 2009 cannot be calculated as authors reported only the mean values. Further, we have added interquartile range for Stolz 2009.
PR (#3)	Results	Critically Ill/Ventilator-Associated Pneumonia Page 26 of full report (Page 55 of 283) i. First paragraph, first line – “increase the power of detecting…” ii. First paragraph, the interpretation of 95% CI is not correct and remove them.	Comment i: We changed “probability:” to “power”, as suggested. Comment ii: The text with the incorrect interpretation of the 95% CI has been deleted..

Commentator & Affiliation	Section	Comment	Response
PR (#3)	Results	Critically III/Ventilator-Associated Pneumonia The grade of evidence, esp. the item of "precision" could be assisted by quantitative synthesis. For example, for ICU LOS, (actually only four studies provided evidence and table 7 says 5 studies), two studies reported significant difference and two studies reported non-significant difference – how do the investigators conclude the evidence to be precise in this case (Table 7)? This also applies to studies of other patient populations.	Clarification has been made in the final report. Four studies reported ICU LOS (Table 6), but the Nobre study was excluded from the meta-analysis because it did not report sufficient data for calculations (Figure 6).
PR (#3)	Results	Procalcitonin-Guided antibiotic intensification Table 10, Jensen P-value could be calculated.	95% CI for Jensen 2011 cannot be calculated as authors reported only the median values with Inter Quartile Range.
PR (#3)	Results	Procalcitonin-Guided antibiotic intensification Table 12, 95% CI for Jensen could be calculated. Please also check this for tables for the rest of the report.	95% CI for Jensen 2011 cannot be calculated as authors reported only the median values with Inter Quartile Range.
PR (#3)	Results	Respiratory Tract Infections Be careful of reporting insignificant results here and in other places of the results. For example. antibiotic use (page 36 in the full report and page 65 fo 283): "Only one study reported an absolute increase in antibiotic duration of 0.1 days with a relative increase of 1.3 percent" -- it would be more appropriate to say "No difference" instead of "absolute increase". Another example "only one study reported a reduction in mortality with procalcitonin-guided therapy by -0.1 percent".	We believe our language is more exact and prefer to retain it. Whether a small between-group difference should be considered no difference depends on statistical power and clinical significance.
PR (#3)	Results	Respiratory Tract Infections Assess the appropriateness of meta-analysis for various outcomes in this population.	Meta-analyses have been added for antibiotic duration, antibiotic prescription rate, short-term mortality and ICU admission.
PR (#3)	Results	Respiratory Tract Infections Another example of grading evidence: ICU admission – none of the studies reported significant results and evidence is precise?? A metaanalysis could help clarify. Also 28-day mortality?	We agree that meta-analysis assists in assessing consistency and precisions. As noted above, we performed a meta-analysis on ICU admission. We also performed a meta-analysis of ≤ 6 week mortality.

Commentator & Affiliation	Section	Comment	Response
PR (#6)	Results	Page 17, Figure 3. PRISMA diagram for identified gray literature. In my experience, I have read or seen a fair number of PRISMA diagrams. However, this particular one somewhat continues to elude me as to where the eventual 4 studies that were included in the report came from. Here's my take on it (and if incorrect, the authors should consider how they could better describe the information). If we look down the full text review column, the two included studies were from the 29 ClinicalTrials.gov site and the 1 included study below it was from the 33 ICAAC, IDSA, ACCP, PAS. Had to work a little bit to figure this out. But I probably have it right? However of the 67 Thermo Fisher Scientific citations where 65 were excluded, what is the interpretation of the box in the full text review column? Two were included, then one of those was excluded because of meeting exclusion criterion? Then why would it be included in the first place?	The reviewer interpretation of the PRISMA for gray literature is correct. We feel that the gray PRISMA diagram in the report clearly describes the process we followed for gray literature search. Further, we excluded one study during full text review because it met the exclusion criterion. We were unable to exclude this study during the stage of abstract review because we were unable to judge whether it met the exclusion criteria or not.
PR (#6)	Results	Page 18, Line 34 - 57: Somewhere in this paragraph or in a short paragraph that follows, I would provide brief descriptive text about the categories of PCT testing methods used in the RCT's. This information is in the abstraction and evidence tables in Appendix C as well as generally referred to in the introduction. However, the introduction states there are 16 studies that used the quantitative Brahams Kryptor, one that used a semi-quantitative version of this assay and one that didn't record the method for PCT measurement. However, when I poured through Appendix C (abstraction and evidence tables), I seemed to come up with 16 studies that used the quantitative Brahams Krpytor and 2 that used the semi-quantitative version? It would be helpful if this was generally stated in the section of the report summarizing characteristics of the 18 RCT's. That is, which two studies are not like the other 16? Also, if it was relevant for these two studies (Manzano et al and Svoboda et al), mention if they were the sole study in any of the populations that had only one study in it.	Types of PCT assays used in the studies were reviewed, and clarifications were made in the final CER. Sixteen studies used a BRAHMS® quantitative procalcitonin assay, and the remaining two studies (Manzano and Svboda) used the BRAHMS® semi-quantitative assay.

Commentator & Affiliation	Section	Comment	Response
PR (#6)	Results	Page 46, Line 17: In the above regard, the one poor study in the fever of unknown source (children aged 1 - 36 months) is the Manzano et al study that employed a semi-quantitative PCT method. Don't know if that is relevant, but I do feel it is worth stating in the study characteristics.	We have made necessary edits to clarify this information.
PR (#6)	Results	Overall, I had no additional concerns, comments or questions on the results section. The level of detail is appropriate and the characteristics of studies are clear. The figures, tables and appendices were otherwise effective and adequate. I am not aware of other studies that should have been included and don't feel that any of the 18 RCT's included ahouls have been excluded. I thought the information was laid out in a clear and comprehensible fashion which was significantly helped by structuring the CER around the five patient populations that were studied in these 18 RCT's.	No response necessary
PR (#7)	Results	Results: Very exhaustive methodology but limited by the quality of studies.	No response necessary
PR (#2)	Discussion Conclusion	Discussion/ Conclusion: The discussion, summary of findings, discussion of limitations and items for further study are all well presented. The specific elements for future study present a very clear road map for future research.	No response necessary
PR (#6)	Discussion Conclusion	This summary and discussion is a clear and solid section. Implications of the major findings are clearly stated as well as the features that differentiate this systematic review from the four others that have been done previously. Table 31 summarizing the scope of all five systematic reviews was helpful and a good illustration.	No response necessary
PR (#6)	Discussion Conclusion	Pages 63 - 66: While the research gaps enumerated, described and discussed are clear and appropriate, I can't say that each of these is easily translated into new research. The first two research gaps are straightforward although not easy studies to design. The last three research gaps are appropriate given the information in this report although challenging to design studies around. More and more institutions are implementing antibiotic stewardship programs. So there is no doubt that comparing that approach to use of PCT to guide antibiotic therapy would be an important RCT to conduct. But difficult to design and implement in the real world.	No response necessary

Commentator & Affiliation	Section	Comment	Response
PR (#7)	Discussion Conclusion	The bottom line is that how can I use procalcitonin levels to manage my patient who was in septic shock after 14 days of antibiotic therapy who is on dialysis, mechanical ventilation after a hospital acquired pneumonia and line sepsis.	Recommendations for clinical management of specific patient scenarios is beyond the scope of this CER.

Commentator & Affiliation	Section	Comment	Response
PR #1	General Comments	Our primary critique is that the authors conclude rather definitively (high evidence) that PCT guidance reduces antibiotic duration in the ICU and initiation/duration in LRTI. We believe that the current evidence does not support such a definitive conclusion as the findings may not be applicable to US practice. In addition, evidence for effectiveness in the ICU is considerably weaker than that for ED LRTI. None of the trials were conducted in the U.S. In addition to the obvious between-country generalizability issue, closer reading of the largest trials methods shows PCT intervention implementation strategies that do not appear realistic in the U.S. For example, in the 2004 ProRESP and 2006 ProCAP studies, algorithm advice was often directly delivered by a senior investigator (not listed in papers, but publicly presented in lectures by this research group) who potentially influenced ED staff. In the 2009 ProHOSP study, "enforcement", as per the published paper, was accomplished by requiring the treating physician to follow Web-based instructions before registering their patient into the study. Further, in order to overrule the PCT algorithm, the treating physician had to first consult the study's coordinating center. In routine US clinical practice, clinicians would not tolerate such interference in their clinical decision-making from study personnel. Thus, it is uncertain if the impressive results of these past studies could be replicated in the US. The first author (P. Scheutz) of the ProHOSP study has also publicly stated the need for a large US confirmatory trial. We suggest concluding that evidence from non-US studies is strong, but generalizability to US practice is unclear given between-country practice differences and research methods not replicable in the U.S., and that future research would include attempting to replicate these results in the US.	We disagree that the evidence for critically-ill is weaker than that for lower respiratory tract infections. U.S. clinicians respond to clinical practice guidelines. The strength of evidence is high to conclude that procalcitonin guidance results in a shorter duration of antibiotic therapy. We have added meta-analyses on antibiotic duration for both critically ill/ventilator-associated pneumonia (VAP) and respiratory tract infections (RTIs). For critically ill/VAP, we had sufficient information to perform meta-analysis using 3 of 5 studies that reported antibiotic duration, showing a significant advantage favoring procalcitonin guidance by 2.05 days. The 2 excluded studies reported effects larger than any of the included studies, so if these 2 studies had been included, the pooled effect estimate would have been even more favorable for procalcitonin guidance. For RTI, 4 of 7 studies reporting on this outcome provided sufficient information to be included in meta-analysis, yielding a significant effect favoring procalcitonin guidance by 2.35 days. Of the 3 excluded studies, the largest study enrolled 3 times as many participants as the combined samples of the 2 smaller studies. The large study reported an effect larger that the pooled effect (-3.0 days), while the 2 smaller studies reported smaller effects (-2.0 days, -1.7 days). Using a conservative assumption of imputing standard deviations as high as the highest value among included studies would produce a pooled estimate similar to the original one and a narrower 95% confidence interval. Regarding the comment that the findings mare not be applicable to US practice, this does not undermine the strength of evidence rating, but is relevant to statements of applicability. This is mostly an operational question. If a PCT-guided algorithm reduces antibiotic usage in a study setting, whether or not an institution will be able to implement the same algorithm and obtain similar benefit will be institution-specific and is not the focus of this review. If however the institution could achieve adherence rates similar to these published studies, they could expect similar results, regardless of whether the institution is located in the US or Europe. A statement addressing this point has been added to the Summary and Discussion chapter section on applicability. Adherence to algorithms varied in studies included in this review, and we tabulated the rates of non-adherence to study algorithms.

Commentator & Affiliation	Section	Comment	Response
PR #1	General Concerns	Our second main concern is safety. The authors correctly note issues with non-inferiority margins and appropriately note that definitively determining a mortality difference would require a very large sample size. However, their main conclusive statements don't include these concerns and their research gaps also do not include these concerns. The authors' meta-analysis of the 5 ICU studies is also a bit falsely reassuring, given that of the 5 studies all except one (PRORATA) were quite small (27 – 111 patients each). And, 90-day mortality, not 28-day mortality, is more and more becoming the benchmark for ICU studies. PRORATA's finding of a slightly higher 90day mortality is particularly concerning, esp as there is reasonable biologic plausibility for harm (see: Chastre 8 vs 14 days abx for VAP - short course did worse for certain infections).We suggest concluding that although PCT guidance appears safe, true safety is unknown, and future research should address this important issue before widespread implementation.	Concerns about non-inferiority margins are addressed as a methodologic weakness highlighted in the Summary and Discussion chapter. We note uncertainty about the choice of noninferiority margin in clinical trials. Summaries of the main findings in the Summary and Discussion chapter also include statements about noninferiority margins and precision of effects on mortality. One of the key reasons to do a meta-analysis is to overcome the imprecision in smaller studies. We disagree that 90 day mortality is the benchmark for ICU studies. The Bouadma study looked at 60 day, not 90 day, mortality. No studies reported 90 -day mortality. It is recommended that early mortality (28 day or in hospital) be the primary endpoint for sepsis studies. Mortality at later endpoints is more often related to underlying co-morbidities. Due to the uncertainty of the appropriate non-inferiority margin, the strength of evidence for mortality in critically ill patients was downgraded to low, but we stand by our conclusions that use of procalcitonin guidance does not increase mortality and morbidity among the critically ill/VAP and RTI.

Commentator & Affiliation	Section	Comment	Response
PR #1	General Comments	The Jensen paper had many flaws. Most importantly, as deployed in their study design, PCT had little chance of affecting outcome. Why would a physician alter his/her treatment plan based on a high PCT obtained early in the pt's course + management plan? And why would one expect any benefit from any alteration in management so early in a patient's course? I.e., too early to declare a high PCT a treatment failure. The authors also credit the Jensen paper for explicitly detailing the intervention. However, when we read a submitted version of this paper we found their description of the intervention vague and found many discrepancies with their previous trial methodology paper. As a minor point the authors in the ES only discuss broadening abx coverage in this study, but not the extra source-hunt testing.	The investigators were using persistently high ("alert" procalcitonin) as an indicator that the patient may be on inappropriate initial antibiotic therapy, i.e., the initial regimen does not cover the infecting pathogen. Unless this is rectified quickly very early in the course of therapy by expanding coverage, mortality could be significantly increased. The same concept is true with finding an underlying infectious process (i.e., an undrained abscess). Regarding confidence in the methods used in the Jensen trial, the protocol for this study was published in BMJ and it outlined the microbiologic and radiographic studies to be ordered if there was an "alert" PCT value and the type of alteration, expansion of antibiotic therapy that should be done in response to the alert. The authors did not reproduce the algorithms/guidelines for expanding ABT therapy in the final paper, but this was the same as that published in BMJ. In fact, 703 patients were already enrolled in PASS at the time the protocol was published in BMJ. In this study, however, there was no difference in the utilization of radiographic studies in the PCT-guided arm. There were differences in antibiotic utilization as a direct result of the intensification protocol. It is this antibiotic intensification that actually had a negative effect on outcomes. The ES only discussed antibiotics since this was the major theme of the entire systematic review. Only 2 studies (Jensen, Svoboda) looked at intensifying the radiographic evaluation, so it was not possible to evaluate this.
PR #1	General Comments	Authors should note that older studies, and thus meta-analyses, are limited by using older, less sensitive PCT assays. All the recent trials and observational studies use the high-sensitivity assay.	The PCT cut-offs for the algorithms were similar between studies and the sensitivity of the assay is not relevant. Below-cutoff low concentrations of PCT detected with a high-sensitivity assay would not alter the management based on the algorithm.
PR #1	General Comments	Table of Contents does not match content. Eg, Exec Sum is listed as 1 page, but it actually goes on for 17 pages.	Revised as requested.
PR #1	General Comments	Executive Summary is 17 pages. Full report is ~54 pages. Should Exec Sum be shorter?	No response necessary
PR #1	General Comments	Dr. Schuetz did publish one paper that somewhat addresses long-term impact of using PCT guidance in the real world. http://www.ncbi.nlm.nih.gov/pubmed/20039090	This was a noncomparative observational study, and the study did not meet inclusion criteria
PR #1	General Comments	Antibiotic stewardship programs are excellent for > Day 1 care (e.g. narrowing abx on ICU Day 3), but we believe less practical and impactful for ED care where quick decisions w/ minimal data are needed.	Antibiotic stewardship programs were not reviewed and were not a focus of this review. Although antibiotic stewardship programs would be a good comparator, there were no RCT's comparing PCT guidance to antibiotic stewardship programs.
PR #1	General Comments	Page 19, line 34 "constituted antibiotic"? typo?	This typographical error was corrected in final draft.

Commentator & Affiliation	Section	Comment	Response
PR #1	General Comments	Ref 52 is not a primary reference to show that 75% of all ambulatory abx are for resp infections.	Change sentence to, "In 2005 to 2006, 54% of ambulatory visits for an acute respiratory tract infection resulted in an antibiotic prescription…" Reference: Grijalva CG, Nuorti JP, Griffin MR. Antibiotic prescription rates for acute respiratory tract infections in US ambulatory settings. JAMA 2009 Aug:302(7):758-66.
PR #2	General Comments	General Comments: This report addresses a clinically relevant set of questions. The target populations and settings are generalizable and the results have immediate meaning for clinical decision making. The key questions and analytic framework are nicely presented.	No response necessary
	General Comments	Clarity and Usability: Extremely well organized. Use of tables and figures is excellent.	No response necessary
PR (#5)	General Comments	This is a thorough and careful review of the use of PCT to guide antimicrobial therapy. I have only one substantial critique which relates to the analysis of PCT to discontinue Abx in the critically ill/VAP population. My concern is that this analysis doesn't sufficiently emphasize the limitations of the studies. Specifically, I am concerned about both the conclusion related to effectiveness (reduces Abx use) and that concerning mortality (mortality not increased). I'll discuss these separately:	Concerns about effectiveness and mortality are addressed below
PR (#5)	General Comments	Strength of evidence was judged to be high that procalcitonin guidance reduces antibiotic usage. I actually agree with this conclusion, including the strength of evidence. But haven't we set the bar awfully low? After all, nearly any protocol that institutes new triggers for discontinuation will, in fact, lead to reduced use. To follow my analogy on our (long ago) conference call, a daily coin toss (heads we stop Abx; tails we continue) will reduce Abx usage. This, in itself, is unremarkable and (more importantly) must be considered in light of the effect on outcomes, which I consider next.	Other interventions, such as use of practice guidelines that dictate the duration of therapy strategies that utilize clinical features for discontinuing ABTs, have been used to decrease the duration of ABT therapy (one study). Guidelines, however, don't individualize therapy based on a marker that appears to reflect a response to therapy that can be monitored. The reduction in antibiotic therapy in these studies of PCT-guided therapy is as high as that seen with any other intervention. Also, discontinuation of antibiotics or duration of therapy are intermediate outcomes, and this report did look at more patient-centered outcomes, such as mortality and antibiotic adverse effects.
PR (#5)	General Comments	There is moderate evidence that procalcitonin-guided antibiotic discontinuation does not increase mortality or morbidity as indicated by ICU length of stay. I have more concern about this conclusion, along two lines. One, is that the threshold for clinical significance (10% change) is provided without rationale and (I think) is way too high. This limitation is raised by the current	We appreciate these issues as unresolved concerns and we'll discuss more in the discussion section. We agree that a 10 % increase in mortality is not acceptable. In fact, no increase in mortality would be acceptable, but it may be difficult, if not impossible, to determine differences in mortality < 10 %. To demonstrate a difference in mortality between PCT-guided therapy

Commentator & Affiliation	Section	Comment	Response
		review, but I'd go a step further. Since the actual decline in Abx use was a modest 20% (roughly), how much can one really expect mortality to rise? For example, assume for the sake of argument that the 20% reduction in use was all inappropriate (meaning Abx were stopped in patients with true susceptible bacterial infection that would progress without treatment). In order for this to cause a 10% rise in mortality in the overall study group, such discontinuation would have to produce a 50% mortality! Remembering that these are patients being monitored intensively, I'd expect that the doctors would attempt to rescue patients having septic deterioration after discontinuation of Abx. It strikes me that a 50% mortality is grossly unrealistic. Thus I think the 10% threshold for mortality is far too high. More important is a limitation not overtly considered in the manuscript. The ability of a study to detect an increase in mortality or morbidity depends on how much Abx overuse there is in the study population. In the subjects studied, we have no way of knowing the magnitude of this. To clarify my point, imagine that none of the study subjects actually has bacterial infection (admittedly an absurd assumption). In this case, any protocol that reduces Abx will be deemed safe, because mortality will not rise. Assume (somewhat less absurdly) that 50% of subjects are being prescribed Abx inappropriately. In this case, the ability of a study to detect a mortality cost will only be half of that in a study in which there is no Abx overprescription. The conclusions in this manuscript implicitly assume that Abx overuse is zero and I don't think this is likely to be true. A sensitivity analysis could explore the impact on conclusions about mortality based on differing assumptions about Abx overuse – I wonder if this might add value and, of course, would lead to less confident assessments about safety. So, after reading this manuscript, would I actually use PCT in my ICU? No. So when I look at the conclusions (as listed above), especially when given confidence grades of high and moderate, I get uncomfortable. To me, the benefits seem pretty modest (even if real) and the safety uncertain, even before taking cost into account. In my book, PCT is not ready for prime time.	and standard therapy and have a confidence interval that is narrow would require a large sample size; however, the complexity of critically ill patients and variability in mortality of sepsis patients limits researchers' ability to enroll large numbers of critically ill patients in such studies. Thus, an appropriately powered study with thousands of septic patients enrolled is not likely to be feasible. The IDSA recommends a mortality noninferiority margin of 10%.Our discussion of the low strength of evidence rating for mortality in the critically ill/VAP population now raises uncertainty about whether the 10% margin might be too high. The studies used well-accepted criteria for the diagnosis of sepsis and VAP. Although the 10 % margin for mortality for the non-inferiority study by Bouadma is in line with recent guidelines from the IDSA for pneumonia and has been used in other sepsis trials, there is still debate as to what the appropriate mortality noninferiority margin should be. While the benefit, and the urgency, of antibiotic stewardship is recognized, the present evidence does not report on important outcomes such as ABT adverse outcomes, resistance, or super-infection which are necessary to define the acceptable confidence interval on mortality. The salience of the upper limit of the mortality confidence interval depends on the clinical significance of any beneficial outcomes attributable to PCT-guided ABT use. The idea that antibiotic use was inappropriate (no infection) in these studies was reduced as much as possible. In fact, if the patients met entry criteria, even in the absence of infection, ABT therapy is still appropriate. There is no better gold standard for the need to initiate ABT therapy for sepsis than the clinical criteria for the diagnosis of sepsis. Some patients may be put on inappropriate therapy, but there is no way to know how many. It is analogous to having to remove a certain percent of normal appendices in order not to miss a case of acute appendicitis. Morbidity and mortality were unaffected by the reduction in antibiotic use, but most individual studies were not adequately powered to show a difference in mortality. Only the study by Bouadma, the largest study to date, did a power analysis and pre-defined a margin for non-inferiority for 28 day and 60 mortality, and in this study PCT-guided therapy met the non-inferiority margin. Meta-analysis was performed looking at mortality across all 5 studies, and the 95 %

Commentator & Affiliation	Section	Comment	Response
			confidence interval for the difference between PCT-guided therapy and standard care was between was -6 % to 5 %. The Bouadma study predominated in the meta-analysis. We did not include 60-day mortality in our analysis, since this was only reported in the Bouadma study, and there were letters to the editor about this study concerning the increased 60-day mortality in the PCT-guided therapy group. This late mortality, however, is more likely related to the underlying co-morbidities. Interestingly, although the PCT and control groups in this study were similar with respect to individual co-morbidities, there were more total co-morbidities reported for patients in the PCT arm. Furthermore, the authors reviewed the cause of death for patients who died between 29-60 days, and death was due to relapse of infection from a shortened duration of antibiotic therapy. The overuse of antibiotics is not taken into account. A sensitivity analysis that takes antibiotic misuse into account may help but studies do not report this variable so it cannot be performed. We did not make any assumptions about the degree of antibiotic misuse or overuse.

Commentator & Affiliation	Section	Comment	Response
PR (#6)	General Comments	General Comments: This CER on procalcitonin (PCT)-guided antibiotic therapy is one of the more comprehensive and complete systematic reviews that I've read and/or reviewed. This draft is better than most that I've previously reviewed although there are a handful of issues and questions identified in my online review. This draft is still not a final product. Given the option, I would have preferred a rating of very good for this draft as opposed to good or superior. Addressing the clinical utility of biomarkers in sepsis, this CER represents a very focused target of that initial proposal by looking specifically at PCT-guided antibiotic therapy. But this focus is clinically appropriate and the findings of this CER clinically meaningful. The interests regarding PCT that my colleagues and I have, and are aware of, by those in laboratory medicine as well as in healthcare overall are significantly met by the findings on using PCT to guide antibiotic therapy in this CER. There is certainly more work to be done. However, this CER provides the broadest and most comprehensive assessment, to date, of PCT utility for guiding antibiotic therapy. The target population (s) and audience (s)are explicitly defined. The key questions are appropriate and explicitly stated. It was a good decision to organize the results of the final draft CER based on the five patient populations studied and reported on in the 18 RCT's/publications that made the final cut for inclusion in this CER.	No response necessary
	General Comments	Clarity and Usability: Overall, this report is well structured and organized. The main points are clearly presented. Professional associations, such as the ones that I'm active in, should be able to use the findings of this report as foundations for clinical scientific programming and potential development of clinical practice guidelines. The findings, given the strength of the evidence as well as the quality of the studies, for the two patient populations of critically ill/VAP and RTI patients, should be very useful in guiding future clinical practice and, potentially, health policy decisions.	No response necessary

Commentator & Affiliation	Section	Comment	Response
PR (#7)	General Comments	General Comments: Overall, the most exhausting report I have ever reviewed. This is not for the average clinician consumption so the purpose and audience is unclear.	No response necessary. AHRQ reports have not contained a section stating the intended audience. While the review tends to emphasize specialized discussions of methodology and clinical content, the synthesis and conclusions are relevant to a broad audience of patients, clinicians and policymakers. Translational documents created in collaboration with the Eisenberg Center are tailored specifically to these different audiences, but the CER is written for those familiar with systematic review and the clinical content area.
PR (#7)		Clarity and Usability: Dellinger RP, Levy MM, Carlet JM, et al. Surviving Sepsis Campaign: International guidelines for management of severe sepsis and septic shock: 2008. Crit Care Med. Jan 2008;36(1):296-327.	We will review and incorporate historical importance into Background section
PR (#6)	Appendix	Appendix C, Page C-71, Evidence Table 1R: In the assay type, the PCT-Q, Brahams is followed by the phrase 'good correlation with PCT LUMI.' What is meant by LUMI? Luminescence method? Ambiguous and vague.	Yes, Lumi signifies luminescence.