

## *Comparative Effectiveness Research Review Disposition of Comments Report*

### **Research Review Title: Benefits and Harms of Routine Preoperative Testing: Comparative Effectiveness**

Draft review available for public comment from August 15, 2013 to September 10, 2013

Research Review Citation: Balk EM, Earley A, Hadar N, Shah N, Trikalinos TA. Benefits and Harms of Routine Preoperative Testing: Comparative Effectiveness. Comparative Effectiveness Review No. 130. (Prepared by Brown Evidence-based Practice Center under Contract No. 290-2012-0012-I.) AHRQ Publication No. 14-EHC009-EF. Rockville, MD: Agency for Healthcare Research and Quality; January 2014. [www.effectivehealthcare.ahrq.gov/reports/final.cfm](http://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.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Executive Summary	While the report is fair, the absence of useful findings fails to direct MD decision-making.	We agree that the evidence is highly limited.
Peer Reviewer #1	Executive Summary	The clinical population was defined by the literature available or, rather, the absence of literature.	No response.
Peer Reviewer #1	Executive Summary	The key questions are physicians' key questions and are well-stated.	Thank you.
Peer Reviewer #2	Executive Summary	Key Questions excellent and explicit.	Thank you.
Peer Reviewer #2	Executive Summary	Audience well defined	Thank you.
Peer Reviewer #4	Executive Summary	The value of routine preoperative testing is very clinically meaningful, as well as extremely relevant to test utilization and costs.	No response.
Peer Reviewer #4	Executive Summary	Testing that has no added value can be in fact harmful, and draining to scare resources.	We agree, and thus have written: During the past three decades, routine preoperative testing has been challenged by several academic publications with concerns about the sizable cost of testing, overtesting and the consequences of false positive tests (leading to unnecessary workups and treatments), and the unknown benefit to patients. <sup>3-8</sup> In addition to increasing the cost of surgical care, <sup>2</sup> nonselective preoperative testing may result in false positive or borderline results (in the absence of clinical indication) which require further investigation. Additional investigation may cause unnecessary psychological and economic burdens, postponement of surgery, and even morbidity and mortality as a result of unnecessary evaluation (e.g., complications due to unnecessary biopsies performed to follow up false positive laboratory tests). <sup>2</sup>
Peer Reviewer #4	Executive Summary	The target population is explicitly defined, although the heterogeneity made conclusions difficult.	Thank you.
Peer Reviewer #4	Executive Summary	Key questions are explicitly stated.	Thank you.

Source: <http://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=1846>

Published Online: January 29, 2014

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #4	Executive Summary	Although perioperative clinical outcomes were included in the questions, they were not part of the assessment. (ES3-24) It would have been of great added value if these assessments could have been included, i.e., complications when testing vs non-testing.	We have clarified the unclear text: The review focuses on the direct evidence of the comparative value of routine preoperative testing versus not testing (or other protocols for testing). This evidence is derived primarily from studies that directly compare testing protocols. These are the only studies that can demonstrate whether uniformly testing an unselected population prior to surgery leads to better outcomes for those patients. We also included cohort studies that report rates of “process outcomes” (rates of surgery cancellation, changes to planned surgery or anesthesia, etc.) only for patients being tested since the rate of procedure delay and cancellation, etc., due to testing is, by definition, zero in patients who do not undergo testing.
Peer Reviewer #5	Executive Summary	The report identifies the key questions and unanswered issues.	Thank you.
Peer Reviewer #5	Executive Summary	It identifies that the evidence in this area is poor.	Thank you.
Peer Reviewer #5	Executive Summary	Unfortunately, the incremental value of the information is small.	We agree
Peer Reviewer #6	Executive Summary	Pg ES-19; line 37: The sentence about “complications and death” being more common in ad hoc testing gps and how this pertains to difficulty in extrapolating cataract findings to other populations is not clear as written. In the full article it is more clearly written (in conclusion). I suggest re-writing this in the ES section.	To improve clarity, we have rewritten this sentence in the Conclusions: Based on high risk of bias studies there is a possibility that complications and deaths occurred more commonly among patients undergoing <i>ad hoc</i> , as opposed to routine or per protocol testing. This raises a caution against extrapolating the cataract findings to other surgeries and populations who may be at higher risk of complications due to the nature of the procedures and the patients underlying illnesses and comorbidities.
Peer Reviewer #6	Executive Summary	There are quite a lot of typos throughout the article. Here are a few:	Thank you. We have fixed these.
Peer Reviewer #6	Executive Summary	ES-1; line 39: postoperative is used instead of preoperative.	Fixed
Peer Reviewer #6	Executive Summary	Urinalysis misspelled in tables 5, 8 and 10.	Fixed
Peer Reviewer #6	Executive Summary	Cancellation misspelled pg 38; line 22.	Fixed throughout
Peer Reviewer #6	Executive Summary	Typos on pg 25; line 21. pg 16; line 44. pg 55; line 49. pg 50; lines 13, 14 & 54. pg 59; line 24.	Thank you.
Peer Reviewer #7	Executive Summary	The report is clinically meaningful, as preoperative testing is an important issue, especially as cost-containment strategies, quality of care issues, and outcomes gain importance in medical care delivery.	Thank you.

Source: <http://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=1846>

Published Online: January 29, 2014

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #7	Executive Summary	I saw a few grammatical and spelling issues:	We have fixed the errors we found.
Peer Reviewer #7	Executive Summary	ES-1, line 40: "postoperative" should be "preoperative"	Thank you.
Peer Reviewer #7	Executive Summary	ES-1, line 40: "define" is missing subject?	Thank you.
Peer Reviewer #7	Executive Summary	ES-3, line 39: "systemically" should be "systematically"	Thank you.
Peer Reviewer #7	Executive Summary	ES-10, line 43: "insufficient for" is missing a word	Thank you.
Peer Reviewer #8	Executive Summary	The topic was nominated by three professional medical associations.	No response.
Peer Reviewer #8	Executive Summary	The key questions follow the typical way that the questions are asked in a CER and are generally appropriate.	Thank you
Peer Reviewer #8	Executive Summary	One point that could be more clarified is the intervention of interest. The KQ implied that the intervention of interest is only routine preoperative testing, the method section says the intervention of interest is routine or per protocol testing and the comparative results are routine or per protocol vs. ad hoc testing.	We have added "or per protocol" to the KQs.
Peer Reviewer #9	Executive Summary	The body of work is terrific.	Thank you!
Peer Reviewer #9	Executive Summary	Unfortunately it [the body of work] is not as clinically meaningful as I had hoped.	We agree.
Peer Reviewer #9	Executive Summary	Though it may stop overuse in some preop settings, it is not a "game changer":.	Probably true.
Peer Reviewer #9	Executive Summary	The calling out of gaps in the evidence will be helpful to those in a position of scholarship/research, but his will impact the day-to-day practitioner less so.	We agree
Peer Reviewer #1	Introduction	Shorter is better but this is OK.	There is much material to cover. Thank you.
Peer Reviewer #2	Introduction	Well-written and inclusive	Thank you.
Peer Reviewer #3	Introduction	A more detailed and proper historical perspective and literature review would be helpful, allowing the readers to understand the issues laid to rest and providing proper perspective on current unsolved problems. For example, including studies exploring the value of chest xray for routine pre-op evaluation, but not necessarily pointing out that in populations characterized by a high prevalence of smoking (e.g. veterans in the 1950s and 1960s), the pre-test probably of missing a small cancer was not negligible, even if the basic CXR was not very capable of finding it.	We decided that it is important to keep this report focused on routine and per protocol preoperative testing and to not expand into a review or discussion about routine testing in general. Though, based on your suggestion, we have added a statement to the Introduction about how preoperative testing (like all routine testing) will find abnormal test results that will lead to new diagnoses (such as previously undetected lung cancer).

Source: <http://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=1846>

Published Online: January 29, 2014

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #4	Introduction	The question of the value of preoperative testing is clearly stated in the introduction.	Thank you.
Peer Reviewer #4	Introduction	Many hospitals have instituted protocols for this testing, but few studies have shown evidence for its value.	No response.
Peer Reviewer #5	Introduction	No comments	No response.
Peer Reviewer #6	Introduction	On pg 1 (line 44): you fail to include the 2012 Cochrane review of testing in cataract surgery. Was this because the Cochrane paper came out after this review was started?	We have added the Cochrane review (which as surmised was not available when the CER was begun).
Peer Reviewer #6	Introduction	The description about patient and procedure heterogeneity is particularly well-written and important to emphasize.	Thank you.
Peer Reviewer #7	Introduction	No issues	No response.
Peer Reviewer #8	Introduction	Page 7, lines 49-51 "An example is testing should be done if the prevalence of an abnormal test is sufficiently low that a sensitive test would yield more false than true positive results." -- Confusing sentence?	This and the preceding sentence about theoretical constructs have been removed.
Peer Reviewer #8	Introduction	Page 8, the labeling of bullets for KQ1 and KQ2 was out of order (it was correct in the ES).	Fixed
Peer Reviewer #8	Introduction	The discussion on patient, procedure and setting heterogeneity is very helpful and provides a good context to understand any results.	Thank you
Peer Reviewer #8	Introduction	The subsection on "preoperative tests" seems to be repetitive and could be incorporated into the first part of the introduction.	We agree that it is somewhat repetitive but we think it is important to keep to clarify the topic.
Peer Reviewer #8	Introduction	In the section of 'assessing the clinical utility', there was the discussion of direct vs. indirect effect and in the subsequent section of "statement of work", there was the discussion of "direct evidence" and "indirect evidence" – seems to be a little confusing in beginning and it helps to add some clarification of distinction.	We rephrased some of the description of direct effects and added a parenthetical statement describing defining direct evidence:  Preoperative testing can have a direct impact only on some outcomes of interest, including emotional and cognitive changes in the patient conferred by testing and its results, any harms associated with the testing procedure (e.g., pain, hemorrhage, or bruising from a blood draw, exposure to ionizing radiation from imaging tests), and costs to the patient (in the form of time spent or copayments) or other types of resource utilization.  direct evidence (evidence regarding actual changes in patient outcomes and management)
Peer Reviewer #9	Introduction	This was good and the summary was great	Thank you.
Peer Reviewer #1	Methods	Criteria are justifiable but, perhaps, not meaningful in the context of available literature.	The eligibility criteria were not based on the available evidence.

Source: <http://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=1846>

Published Online: January 29, 2014

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #2	Methods	Appropriate methods.	Thank you.
Peer Reviewer #3	Methods	As noted above in a different context, it seems that the low risk patients occupy too large a component of the analysis, and there has not been a very concerted effort to look at places where the information might be obtained in higher risk patients.	The same eligibility criteria were used for low and high risk patient studies. Unfortunately, the data available for high risk patients was even more limited than for low risk patients.
Peer Reviewer #3	Methods	The dismissal of reports (page ES3) that don't study prevalence and impact in a direct way seems rigorous, but it also then limits anything the authors could conclude other than to editorialize on the lack of well conducted studies.	We kept the review within the scope of comparative effectiveness of actual testing, not expanding the review to studies that require assumptions be made about how clinicians would respond to testing
Peer Reviewer #4	Methods	The inclusion/exclusion criteria for interventions of interest were overall justifiable, however, the exclusion of patient factors may have effected the findings. (ES-5-32). Not coagulation testing, for example, would not be a choice for a patient with a bleeding disorder history.	The CER was focused on preoperative testing, in contrast with patient history etc. In this section, we describe how patient symptoms are an indication for testing.
Peer Reviewer #4	Methods	Search strategies are clearly stated, logical, and yielded valid results.	Thank you.
Peer Reviewer #4	Methods	The variability of eligible study designs appears justifiable due to the limited number of comparative quantitative and qualitative studies and since cohort studies were limited to "process" outcomes.	Thank you.
Peer Reviewer #5	Methods	Appropriate criteria	Thank you.
Peer Reviewer #6	Methods	The methods section appears to clearly and accurately outline the process.	Thank you.
Peer Reviewer #6	Methods	Search strategies, statistical methods, definitions appear appropriate.	Thank you
Peer Reviewer #6	Methods	pg 12; line 47: not clear what is meant in this sentence re: "We limited these studies that reported "process"... Please reword. Later in the paper this is addressed and is written more clearly.	We have rewritten the sentence.
Peer Reviewer #6	Methods	I wonder if it isn't important enough to include the AHRQ criteria for the grading of evidence and bias as a supplement in this article rather than simply providing a reference since this factors so importantly in the conclusions.	The criteria used are fully summarized in the Methods section. We believe this is adequate.
Peer Reviewer #7	Methods	No issues	No response.
Peer Reviewer #8	Methods	Interventions of interest: please see Executive Summary comments above	Fixed, as above.
Peer Reviewer #8	Methods	Comparator of interest: it seems that the description mixed true comparators in bigger categories and the subgroup variables (setting, timing etc.).	The different comparators do not relate to subgroup variables, but to subquestions.

Source: <http://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=1846>

Published Online: January 29, 2014

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #8	Methods	Categorization of outcomes is clear.	Thank you.
Peer Reviewer #8	Methods	<p>About the eligible study designs – It is very confusing to call “noncomparative, single group studies in which all study participants had the same testing battery or protocol” as a cohort study.</p> <p>A cohort study has a long established and well-defined meaning in epidemiology and is observational but comparative.</p> <p>The included comparative but non-randomized studies may usually be called cohort studies.</p> <p>For these studies, they all have the Interventions of interest, and may be call “intervention series” or “test series”?</p> <p>The term “intervention series” has been used in some reports of our EPC for similar study design.</p>	<p>We originally called these single group studies, but this was thought to be confusing. The term “cohort study” has commonly been used in EPC CERs and other systematic reviews to describe these studies where everyone received the same intervention (or exposure).</p> <p>The suggested terms are new to us.</p> <p>If another term is preferred, we would opt to return to single group studies, which was the term of choice for a related methods report.</p>
Peer Reviewer #8	Methods	And a related question, for these non- comparative studies, how valid it is to assume that change or cancellation of procedures is due to the testing? Change or cancellation of procedures won't occur otherwise?	Good point. We have made this into an explicit caveat at the start of the Cohort Study Findings results.
Peer Reviewer #8	Methods	Data Synthesis: DL method was used to combine OR; however, RR was used in the results section.	This was a typo in the methods section. We meta-analyzed RRs.
Peer Reviewer #8	Methods	Also, the abstract also mentioned Peto's methods.	This was left over from a prior version and has been omitted. Thank you.
Peer Reviewer #8	Methods	<p>A related note, in this review, the number of studies for MA is small but there is not much heterogeneity, so the DL method won't be very bad.</p> <p>However, the new stata routine metaan offers the profile likelihood method and other methods now.</p> <p>It is probably time to start to use a method which takes the between-study variability into better account.</p>	<p>We reran the analysis with the restricted maximum-likelihood random-effects model (reml) which yielded identical results.</p> <p>We also reran with the profile likelihood random-effects model (pl), which had a negative estimate for tau-squared and thus reverted to a fixed effect model. It gave very similar, slightly wider confidence intervals, which do not affect conclusions: 0.99 (0.81, 1.16).</p> <p>For this review, with only a single meta-analysis, we believe it is better to stick with a meta-analytic approach people are more familiar with.</p>
Peer Reviewer #8	Methods	The authors should be commended to specifically considering MID.	Thank you.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #8	Methods	<p>However, the determination of MID for important vs. unimportant difference used difference criteria for mortality and severe morbidities, which may cause problem for some situations. For example, based on the criteria, if RR = 0.9 (95% CI 0.8, 1.01), it will say there was evidence of no important difference. However, if RR = 0.9 (95% CI 0.82, 0.99), then there is evidence of clinically import difference.</p> <p>If MID is 0, it does not seem to make sense to say the former of evidence of no important difference.</p> <p>Conceptually, it is not clearly how justifiable to use different criteria to determine important vs. unimportant difference. IF MID = 0, and for real life data, it is hard to have exact difference to be zero and given enough sample size, we could always detect this difference – then MID = 0 seem to say that there is always a clinically important difference, some detected in the studies and some not due to small sample size.</p> <p>So there is some logical inconsistency in such criteria. In a way, MID has been used more to refer to the point estimate and when testing equivalence or non-inferiority, the CI limits will be used.</p> <p>The criteria used here mixed these two things. For other outcomes, “To determine that there is evidence of a clinically important difference, the 95 percent CI of the difference had to be fully beyond 0.80 or 1.20 (on the RR scale)”.</p> <p>Well, this means if RR = 0.5 (95% CI, 0.28, 0.89), you still could not say there is a clinically important difference, which does not seem to make sense.</p> <p>Also the point estimate RR = 0.5 indicates a difference larger than 20% (the typical definition of MID).</p> <p>Also, 0.80 and 1.20 is not asymmetric around 1 for 95% CI – a problem to determine no important difference (equivalence). It is helpful to clarify that in some cases, evidence could be insufficient to determine whether or not there is clinically important difference.</p>	<p>The MID system has its flaws and still arbitrarily makes distinctions which may be problematic at the extremes (such as these 2 examples).</p> <p>For critical outcomes (e.g., death), it was agreed that any statistically significant difference was important for this low risk (generally low cost) intervention. However, for other outcomes, we wanted to find a clinically meaningful difference (i.e., 20%). We believe this is a fundamental flaw in the MID concept for the situation where any difference is believed to be important (i.e., MID = 0%).</p> <p>The examples you give are correct. The MID system does, in fact, rely on the 95% CIs, not the point estimate. To a large degree, that is the reason for using the system, to account for the uncertainty involved with large confidence intervals. We have also added that “for the purposes of a comparative effectiveness review, the utility of MID pertains primarily to bodies of evidence for which there is sufficient evidence.”</p>
Peer Reviewer #8	Methods	Page 14, lines 35-36: Test of equivalence does look at 95% CI.	We have removed this sentence, as it is somewhat confusing.
Peer Reviewer #8	Methods	Grading the body of evidence Do the authors intentionally choose to use “study limitations” instead of “study quality”?	Yes. Study limitations is a broader concept than study quality.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #8	Methods	The methods section does not include applicability, which were presented in the results section and discussion. For example, the choice of publication year as a criteria to evaluate applicability should be included in the methods.	We have added a paragraph on applicability to the section Grading the Body of Evidence.
Peer Reviewer #9	Methods	Yes	No response
Peer Reviewer #1	Results	Results section is more than sufficient and represents, including graphically, the available evidence well	Thank you.
Peer Reviewer #2	Results	Did not “overreach” on conclusions	Thank you.
Peer Reviewer #3	Results	For the stated intentions of the study, the tables and presentation of data are appropriate.	Thank you.
Peer Reviewer #4	Results	The amount of detail in the results section is appropriate and adequately supports the discussion and conclusion sections.	Thank you.
Peer Reviewer #4	Results	The characteristics of the various surgeries included are clearly described, as are the definitions for routine, per protocol, and ad hoc testing.	Thank you.
Peer Reviewer #4	Results	These are logical areas for comparison.	Thank you.
Peer Reviewer #4	Results	The key messages are clearly stated.	Thank you.
Peer Reviewer #4	Results	Unfortunately for postoperative complications and deaths among patients undergoing routine or per protocol testing, the heterogeneity and flaws in the studies precluded any confidence in the accuracy or validity of the finding.	We agree.
Peer Reviewer #4	Results	Findings that demonstrated the value of this testing would be applicable and extremely valuable to the clinical management of the patient as well as utilization of resources.	We agree.
Peer Reviewer #5	Results	One significant concern is that they combine the results of the 3 RCT around cataract surgery, but the Schein study would overwhelm (by size) the other two. It would be important to outline this potential.	Thank you for this suggestion. We have added this to the results (p. 20).
Peer Reviewer #6	Results	On pg 48; line 10 it is stated “We identified 54 studies...” I thought it was 52 studies?	Thank you. This has been corrected.
Peer Reviewer #6	Results	I think the key messages are there.	Thank you
Peer Reviewer #6	Results	As mentioned for other sections the wording at times appears excessive and redundant.	The repetition was required for completeness and to fit the report’s style.
Peer Reviewer #6	Results	I’m a bit concerned that only readers who are specifically interested in this topic will be able to “get through it” for the key messages.	We believe the ES and the Discussion are accessible to most readers.

Source: <http://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=1846>

Published Online: January 29, 2014

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #7	Results	No issues	No response.
Peer Reviewer #8	Results	The structure of the results is clear.	Thank you.
Peer Reviewer #8	Results	Some places have more details and could be more summarize (e.g., general surgeries for adults or some findings from the non-comparative studies).	We have shortened some sections as appropriate.
Peer Reviewer #8	Results	The summary information is generally clear and helpful.	Thank you.
Peer Reviewer #8	Results	The summary section for non-comparative studies focused too much on subgroup analysis.	Since the studies are noncomparative and provide little information to address the CER questions, per se, the subgroup analyses are particularly relevant.
Peer Reviewer #8	Results	Page 19, lines 39: "The studies had RRs of 1.00 or 0.97, suggesting no difference in cancellation rates." --- No clinically important difference? Since the authors defined clinically important difference, it helps to be clear about this in the text. (0.97 without 95% CI does not say no difference).	We have added the 95% CIs in.
Peer Reviewer #8	Results	Page 20, lines 11-13: "but the confidence interval was too wide to definitely exclude clinically important difference (RR=0.97; 95% CI 0.79, 1.20). " – does this estimate come from one of the two studies?	This is clearer now with the CIs added into the results text. We have also edited the sentence to: "but the confidence intervals were too wide to definitely exclude clinically important (i.e., more than 20 percent) difference."
Peer Reviewer #8	Results	Also Other than the borderline of 95% CI based on your criteria, given RR = 0.97, the estimates seem to say more for evidence of no difference. (the magnitude of the point estimate should be taken into account in some way).	We concluded that there is high strength of evidence suggestion no difference. With the MID framework, though, we have to give a lot of weight to the CI.
Peer Reviewer #8	Results	Page 24, lines 48-55 No justification of the assumption is provided. Why picking out this outcome to do such an analysis by making assumptions? If using a Poisson model which allows multiple event per person, then RR = 0.56 with exact 95% CI (0.31, 1.01) with P > 0.05.	Thank you. We have changed this to the Poisson model RR.
Peer Reviewer #8	Results	Page 25, lines 13-17, " Larocque et al. also found significantly more total complications in patients undergoing ad hoc testing (13%) than per protocol testing (9.2%; P<0.001 by Chi squared or Fisher's exact tests), which results in an almost statistically significant RR (RR=0.71; 95% CI 0.49, 1.01)." -- If P < 0.001 from the Chi-square test, then it is almost impossible to have an RR with 95% CI across 1. Please double check your results.	We have made it more explicit that their reported P value and our calculated RR are not consistent.

Source: <http://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=1846>

Published Online: January 29, 2014

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #8	Results	Page 25, lines 38-40 The event is rare and it is better to calculate an exact 95% CI (no matter what the study reported). Not easy to calculate an exact 95% CI for RR but since the event is rare, OR could be used along with exact 95% CI. This applies to other similar places of the report, for example, page 31, lines 20-30.	We calculated ORs with exact CIs and found very similar results that do not change conclusions. For clarity and simplicity, we believe it is preferable to use standard calculation methods to avoid introducing unnecessary complexity (e.g., RR used some places exact OR others; also the need to determine when events are “rare”). Examples of the differences include:  Finegan complications RR = 0.43 (95% CI 0.13, 1.40) OR, exact = 0.43 (0.14, 1.47) Larocque return to OR RR = 0.25 (95% CI 0.03, 2.19) OR, exact = 0.25 (0.02, 1.48) Meneghini major complications RR = 2.33 (95% CI 0.43, 12.7) OR, exact = 2.32 (0.44, 10.0)
Peer Reviewer #8	Results	Page 31, lines 33-34 “although the rates were similar, they were significantly different (RR = 1.21; 95% CI 1.08, 1.36), favoring ad hoc testing.” – Probably not accurate to say the rates were similar with a 20% significant difference. This is an example that could be used to think whether there is a MID here.	This is actually a great example of when RR may overstate things. We have reiterated the percentages to clarify: “although the rates were similar (15% vs. 13%)”. For an outcome like minor complications, we think this is an appropriate way to consider this. The reader is free to disagree and rely on the RR. Both are given.
Peer Reviewer #8	Results	Page 35, lines 28 – it helps to say it clear that the higher rate is not significant.	Agree
Peer Reviewer #8	Results	Page 38, again, it is confusing to call non-comparative studies “cohort studies”.	See above
Peer Reviewer #8	Results	It is possible to compare rates from non-comparative vs. comparative studies? Would be helpful if the results are consistent.	Given the heterogeneity, it is not clear that such a comparison is meaningful.
Peer Reviewer #8	Results	The non-comparative studies included both routine and per-protocol testing?	Yes, we did not distinguish for these analyses.
Peer Reviewer #8	Results	Page 38 lines 36-45 pooling rates using fixed effects model: first, such methods should be described in the methods section.	We have added this to the Methods section along with a rationale.
Peer Reviewer #8	Results	Second, given the heterogeneity of studies, it may not be the best way to use a fixed effects model.	As noted, this was done only to provide a rough estimate to compare adverse event rates.
Peer Reviewer #8	Results	I understand that there are situations with quite a few studies with zero events, which limits the choice of models but there are situations that appropriate models, fixed or random, could be used. Consider the profile random effects model, too. Also, fixed effect model could be simple pooling, but does not have to be. Please provide a description of methods in the method section with appropriate justification.	Given the heterogeneity and other issues, we decided to use the simplest approach, since regardless of approach used, the estimate would be rough and not generalizable.

Source: <http://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=1846>

Published Online: January 29, 2014

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #8	Results	Page 45 section "Procedures for which testing did not affect outcomes" -- I would exclude this section (and relevant discussion on this in other sections) completely.	We believe this section is of value. The logic behind this section (no AEs occur, therefore preop testing to prevent AEs is unnecessary) is commonly used and rational.
Peer Reviewer #8	Results	Such discussion treats the point estimate (0) as absolutely accurate without considering the 95% CI, which is not appropriate.	We have added in a caveat to address this good point.
Peer Reviewer #8	Results	Also the authors noted that "However, in no scenario (specific test(s) used prior to the same category of procedures in the same population [adults vs. children]) were there at least two studies that both found no changes in patient management." -- in a CER, it is especially not appropriate to pick out one study and claim that the study did not affect outcome.	This is why this sentence was written. Given the heterogeneity of tests and surgeries, almost all our evidence boils down to single studies.
Peer Reviewer #8	Results	Page 45, lines 43-44 "However, we included this outcome because it was the only outcome that was analyzed by patient subgroup" -- this implies that the outcome is chosen post hoc.	The outcome was chosen a priori, but by necessity the interpretation of the outcome was conducted post hoc.
Peer Reviewer #8	Results	Page 46, lines 7-8 "In summary, these cohort studies confirm a greater impact on management by age, ASA category, and surgery risk." -- by increasing age, ASA category and surgery risk?	Thank you.
Peer Reviewer #8	Results	Page 46, lines 12-13 "In all preoperative testing scenarios for which more than a single study was available (i.e., approaching a sufficient evidence base to form a conclusion) resulted ..." -- (i.e., approaching a sufficient evidence base to form a conclusion) appeared many times in the text. However, it makes much more than one study to have a sufficient evidence base, and even "approaching" could not justify it.	It appears only twice. For better or for worse, having two studies can yield a sufficient evidence base. Actually, a single large well conducted study may be sufficient, though clearly that was not the case for this evidence base.
Peer Reviewer #8	Results	Page 46, lines 21-24, the meaning or the purpose of the sentence is not clear?	The sentence was deleted.
Peer Reviewer #1	Discussion	Several of the physicians I queried indicated that the cataract findings may be more applicable to the general population undergoing minimal to moderate procedures than the report indicates.	This is straying a bit beyond the evidence, but we agree that it is a reasonable conclusion. We have added the following sentence to the discussion where the evidence is summarized: While there is no evidence regarding minimally invasive surgeries similar to cataract surgery, it may be valid to conclude that routine preoperative testing in these other low-risk surgeries would also have no effect.
Peer Reviewer #2	Discussion	Limitations are described in detail.	Thank you.
Peer Reviewer #3	Discussion	The limitations, established by the long list of exclusions as well as the catalogue of difficulties in interpreting studies is, in fact, the contribution of the report.	We agree that the limitations of the evidence, together with the call for future research, is the major contribution of the report.

Source: <http://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=1846>

Published Online: January 29, 2014

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #3	Discussion	Chiefly this report is a statement on the criteria for “good” studies connecting prevalence to influence of abnormal findings in asymptomatic patients.	No response.
Peer Reviewer #4	Discussion	Although the evidence is insufficient to clarify specifically which routinely conducted (or per protocol) test may be of benefit, the value and urgency for further studies is well conveyed.	Thank you.
Peer Reviewer #4	Discussion	Limitations of current evidence are clearly detailed.	Thank you.
Peer Reviewer #4	Discussion	Key findings and strength of evidence are clearly described in the discussion section, including why the evidence was not substantial enough for recommendations.	Thank you.
Peer Reviewer #5	Discussion	A key limitation of the research relates to both the protocols of care and the location of care. One of the key issues of preoperative testing is that more testing may be required in places less able to handle complications. Additionally, the way in which testing is utilized can vary. The authors of the review should comment on these issues.	Thank you for these insights. We have added these concepts to the Discussion.
Peer Reviewer #6	Discussion	I think the implications of major findings and limitations outlined appropriately.	Thank you.
Peer Reviewer #6	Discussion	Pg 56; line 7: Need to reword the statement re: ACP guidelines to be clear that they were to reduce “pulmonary” complications not complications in general. Currently written as “preoperative testing to reduce perioperative complications”. Should be: “preoperative testing to reduce perioperative pulmonary complications”.	Corrected.
Peer Reviewer #7	Discussion	I thought the discussion section for the executive summary was perhaps too wordy and long, with perhaps too much editorializing.	We believe that for this review with limited evidence, the Discussion is where the meat of the matter is. We have read through to try to streamline it some.
Peer Reviewer #8	Discussion	For non-comparative studies, I would not say the studies reported “associations” between testing and clinically outcomes. The typical use of association in observational studies implies comparison and differences among groups.	We have rephrased the sentence to remove the word association
Peer Reviewer #8	Discussion	The major comparative results are summarized well but the basis for statement “The apparent difference in the effect of routine (or per protocol) testing in patients undergoing cataract and general elective surgery is arguably not surprising” is not clear. Some differences should be clear and obvious before the review.	We believe the statement is clear and straightforward. There was no effect in cataract surgery, but differences were found for more major surgeries, even if the evidence was not sufficient.
Peer Reviewer #8	Discussion	Adequate discussion of heterogeneity and confounding.	Thank you
Peer Reviewer #9	Discussion	Yes	No response

Source: <http://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=1846>

Published Online: January 29, 2014

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Conclusion	Straightforward implications: in the absence of medical literature no conclusion may be made except that more studies must be performed.	We agree.
Peer Reviewer #2	Conclusion	Future research section very useful.	Thank you.
Peer Reviewer #3	Conclusion	Also noted is the statement on page ES 13 “In other words, the evidence suggests that in most situations, routine preoperative testing will result in some delay or cancellation of the procedure or some changes to anesthetic management or surgical procedure. Again, whether these changes benefit or harm patients is unknown”. The concern in this statement is the implication that such testing leads to delays and slowdowns in care without obvious benefit, but the data provide no information on benefit. The caution here is that the authors may, inadvertently, be arguing that absence of evidence is in fact evidence of absence. The data do not support the latter conclusion.	We think that the phrase “whether these changes benefit or harm patients is unknown from these data” is neutral and balanced. The sentence in question clearly describes an absence of evidence.
Peer Reviewer #3	Conclusion	This paper may serve as an editorial review, guiding discussion about how future studies might or might not be conducted, in order to determine the circumstances and patient groups in which routine testing might be a benefit.	We agree that the limitations of the evidence, together with the call for future research, is the major contribution of the report.
Peer Reviewer #4	Conclusion	The future research section clearly describes what is needed to make valid recommendations for appropriate preoperative testing and how this can be accomplished.	Thank you.
Peer Reviewer #4	Conclusion	The information includes valuable directives for study design, scenarios, categories, etc. for future research.	Thank you.
Peer Reviewer #4	Conclusion	Except for cataract surgery, the conclusions could be used to inform policy and/or practice decisions, but evidence is not sufficient for the sole basis for decisions.	We agree.
Peer Reviewer #6	Conclusion	The conclusions are straightforward.	Thank you.
Peer Reviewer #6	Conclusion	The conclusions regarding cataract surgery pts is quite specific and should drive practice decisions and policy.	No response.
Peer Reviewer #6	Conclusion	CMS should use this to stop reimbursements for these pts	No response
Peer Reviewer #6	Conclusion	Future research quite clear and on the money.	Thank you.
Peer Reviewer #6	Conclusion	Additionally, the challenge and descriptions for proposals of future studies to address the lack of data is very helpful, specific and clear.	Thank you.
Peer Reviewer #6	Conclusion	This should inform AHRQ/NIH to assist in funding appropriate endeavors.	No response.

Source: <http://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=1846>

Published Online: January 29, 2014

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #8	Conclusion	For future research, it would be more helpful to have a discussion in terms of gaps of evidence, instead of by study design (RCTs vs. Observational studies. Vs. decision models).	After much discussion, we decided that the structure used is clearest. Most reviewers seem to agree.
Peer Reviewer #8	Conclusion	The future research section talks too much about advantages and disadvantages of the study designs themselves.	We believe the future research discussion is appropriate.
Peer Reviewer #8	Conclusion	It is clear that there are tremendous gaps in the evidence and it is impossible to have evidence for every category. So it is important to prioritize what is more important. The report mentioned this but did not do this adequately.	We have taken the approach that it is our role to highlight the gaps but for others to prioritize future research needs.
Peer Reviewer #9	Conclusion	Yes	No response
Peer Reviewer #4	Figures	The tables and figures clearly depict the findings and are easy to understand.	Thank you.
Peer Reviewer #6	Figures	Table 1 has an interesting categorization of surgical risk. I am not intimately familiar with the reference and I do believe the literature is not clear on surgical risk nor are most categories based on sound data. It seems odd that a thyroidectomy would be ranked higher than a tonsillectomy and adenoidectomy. And, that a thyroidectomy would be considered major surgery. What is "endoscopic" prostate resection? Is this a TURP or laparoscopic prostatectomy? If it is a TURP then I don't think it is comparable to a TAH.	We provided this table as an example. We have made it more explicit that this is an example of a system. We agree and have removed "endoscopic" prostate resection (NICE's term) and thyroidectomy.
Peer Reviewer #6	Figures	Figs 4-7 are odd representations of information. Perhaps this is just my lack of familiarity but this doesn't seem to be a very useful or straightforward method of presenting information.	Most reviewers apparently found the figures to be clear. We believe we chose a good method to present the data.
Peer Reviewer #6	Figures	Otherwise the tables present a large amt of info in a very accessible manner.	Thank you.
Peer Reviewer #8	Figures	Figures 4-7 could be improved – using letters does not seem to be clearest way to convey information; overlapping of letters is distracting and no information on 95% CI for each estimate is presented. It would be informative to present the data more like a forest plot.	We think that letters are clearer than symbols and that words or abbreviations would be even busier. It is true that the 95% CIs are not included, but adding them would make the figures a jungle of lines. Given the volume of data, we believe what we have presented is clearest. The details are in the appendix.
Peer Reviewer #8	Figures	I like the tables -- the data are presented clearly and organized.	Thank you.
Peer Reviewer #8	Figures	Page 19, lines 24-29 The description of the text is not consistent with Figure 3. For example, the RRs did not range from 0.70 to 2.0 and for 95% CIs, why picking 0.86 and 1.17?	Thank you. This has been corrected.
Peer Reviewer #9	Figures	I could not view the figures.	No response

Source: <http://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=1846>

Published Online: January 29, 2014

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	General	Was easy to review verbally with other physicians.	Thank you.
Peer Reviewer #2	General	Very thorough	Thank you.
Peer Reviewer #2	General	Well organized, nice summary	Thank you.
Peer Reviewer #2	General	Too bad there's not better evidence!	We agree.
Peer Reviewer #3	General	<p>Four major concerns about this report:</p> <p>1) The report contains information that has been analyzed in detail, and in recent publications that the authors should be aware of. Systematic reviews include: Health Technol Assess. 2012 Dec;16(50):i-xvi, 1-159. What is the value of routinely testing full blood count, electrolytes and urea, and pulmonary function tests before elective surgery in patients with no apparent clinical indication and in subgroups of patients with common comorbidities: a systematic review of the clinical and cost-effective literature. Czoski-Murray C, Lloyd Jones M, McCabe C, Claxton K, Oluboyede Y, Roberts J, Nicholl JP, Rees A, Reilly CS, Young D, Fleming T. doi: 10.1093/bja/aet071. Epub 2013 Apr 11. Effectiveness of non-cardiac preoperative testing in non-cardiac elective surgery: a systematic review. Johansson T, Fritsch G, Flamm M, Hansbauer B, Bachofner N, Mann E, Bock M, Sönnichsen AC. In addition, a recent report from the group in Galvestin in Annals of Surgery (Preoperative laboratory testing in patients undergoing elective, low-risk ambulatory surgery. Benarroch-Gampel J, Sheffield KM, Duncan CB, Brown KM, Han Y, Townsend CM Jr, Riall TS. Ann Surg. 2012 Sep;256(3):518-28. has looked at the problem of testing in low risk patients.</p>	<p>Thank you.</p> <p>The HTA (Czoski-Murray et al.) was narrow in scope and did not include comparative studies (for an unclear reason). We have added it to the Introduction of the main report.</p> <p>The review by Johanseon et al. came out after our search for systematic reviews. Thank you for calling it to our attention. However, the review appears to have included only comparative studies (RCTs) for cataract surgery. The rest of their analyses focus on the predictive value of abnormal test results, which we did not evaluate.</p> <p>The study by Benarroch-Gampel et al. did not meet our eligibility criteria since there is no distinction between routine, per protocol, and ad hoc testing. The study evaluates all preoperative testing. The study also did not have outcomes of interest for this report.</p>
Peer Reviewer #3	General	And the conclusion from all such analyses is that routine testing in low risk patient groups is not helpful.	We agree
Peer Reviewer #3	General	Moreover some of the information is really outdated, e.g., the value of routine testing in patients undergoing cataract surgery.	While many of the studies are old, we do not believe the cataract studies are outdated. The age of the studies is only a minor limitation to the poor data available.
Peer Reviewer #3	General	The authors are including in this review topics long ago laid to rest, thereby diminishing the impact and currency of the report.	The report was designed to be comprehensive so it includes a wide range of surgeries and tests, regardless of prior reviews.

Source: <http://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=1846>

Published Online: January 29, 2014

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #3	General	<p>2. Another issue is that the target populations are not defined in a helpful way.</p> <p>In the present day, it does not seem meaningful to look at routine testing as a problem if the target is all surgical patients, of varying medical complexity, undergoing procedures of varying complexity.</p> <p>As noted by the authors, the unsolved problem is the population of patients at risk: patients who are medically complex, patients who are undergoing technically complex procedures or requiring complex and prolonged anesthetics.</p> <p>Thus their statement is perplexing (ES 2): "Although it has yet to be demonstrated, one could expect that some preoperative tests may be of greater value in predicting and ultimately reducing complications in higher rather than lower risk surgeries." What is perplexing is that low risk patients/procedures have been well shown not to benefit from routine testing to pick up chance abnormalities and so the true target population are patients at higher risk.</p>	<p>Unfortunately, we are limited by the data and the large majority of studies targeted broad range of surgeries.</p> <p>As was the case in prior systematic reviews and as becomes evident from the review findings, the evidence is good for cataract surgery (no effect) but insufficient for all other surgeries, higher and lower risk. We do not agree with the assessment that it has been well shown that routine testing does not benefit low risk patients/procedures. One may conclude this based on the cataract data or from the low likelihood of abnormal tests in low-risk populations, but this is conjecture (or at least an assumption in logic that this CER does not make).</p>
Peer Reviewer #3	General	<p>3. the target audience is not stated and this contributes to the lack of cohesion in the report. The information in this report and analysis is well known to surgeons, anesthesiologists and peri-operative care providers.</p>	<p>We have added the target audience to the Statement of Work. We agree that certain aspects of the review are well known, but this review provides the most comprehensive review of the evidence to date and provides a critique of the evidence and recommendations for future research.</p>
Peer Reviewer #3	General	<p>4. This report lacks the current context of laboratory testing in the context of the modern pre-operation/pre-anesthesia evaluation center.</p> <p>This has been explored generally and in different ways by Bader, Gawande and Makary; and the authors do not provide any sort of perspective on using these environments to defining groups with high pre-test probabilities.</p> <p>This was an important opportunity, not capitalized.</p>	<p>We sought evidence on these centers but failed to find eligible studies about them. This CER lays out and evaluates the clinical evidence. Since the evidence is lacking, there is no opportunity for this review to come to conclusions. We leave that for editorialists, narrative review authors, and other thought leaders in the field.</p>
Peer Reviewer #4	General	<p>The report is well structured and organized and follows a logical succession to conclusions.</p>	<p>Thank you.</p>
Peer Reviewer #4	General	<p>The evidence gaps and future research sections are excellent and very informative concerning is what is critically missing to develop high quality improvements in the value of preoperative testing and what actions can be taken to achieve this.</p>	<p>Thank you.</p>
Peer Reviewer #5	General	<p>Clarity and Usability: Acceptable</p>	<p>Thank you.</p>

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #6	General	I am puzzled as to why the article by Frances Chung, et al. "Elimination of Preoperative Testing in Ambulatory Surgery", Anesth Analg 2009;108:467- is not included in this review. Chung, et al. "Elimination of Preoperative Testing in Ambulatory Surgery", Anesth Analg 2009;108:467- should be included in this review.	Thank you for noting this omission. We went back to our search to see why it was missed and discovered that it was not adequately sensitive. We revised the search and based on that found three new eligible studies, including Chung, which have been included. The additional studies did not change the conclusions.
Peer Reviewer #6	General	A bit wordy and redundant. Would be ideal to shorten, simplify without eliminating important content.	We have read through again and made some revisions.
Peer Reviewer #6	General	Chung, et al. "Elimination of Preoperative Testing in Ambulatory Surgery", Anesth Analg 2009;108:467- is not included in this review.	See above.
Peer Reviewer #6	General	Cochrane review on cataract testing published 2012 needs to be included or at least a statement needs to be included as to why it is not discussed with the other 3 major "reviews" (ACP, ASA and ACC/AHA)	See above
Peer Reviewer #6	General	The report is well structured and organized and the main points clearly presented. But, it is a bit wordy.	Thank you.
Peer Reviewer #6	General	I'm not sure it can be shortened without losing the important information though it does appear to be somewhat redundant in certain sections and I believe it can be shortened and simplified making it more accessible to the average reader.	Thank you.
Peer Reviewer #7	General	I thought the authors and study committee did a good job overall with designing and implementing the review.	Thank you.
Peer Reviewer #7	General	Clarity and Usability: No issues	No response.
Peer Reviewer #8	General	The report is well structured and organized.	Thank you
Peer Reviewer #8	General	The main points in the conclusion are clear though they are more about the gap of evidence.	We agree.
Peer Reviewer #8	General	It helps to say what better evidence is most needed to be developed.	As noted above, we have taken the approach that it is our role to highlight the gaps but for others to prioritize future research needs.

Source: <http://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=1846>

Published Online: January 29, 2014

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #8	General	As a specific comment, “Nevertheless, the suggestion that complications and deaths occurred more commonly among patients undergoing ad hoc testing raises a caution against extrapolating the cataract findings to other surgeries and populations who may be at higher risk of complications due to the nature of the procedures and the patients underlying illnesses and comorbidities.” -- Other than that this is a very long sentence, as mentioned earlier, the inappropriateness of extrapolation is probably more due to the differences in different type of surgeries (known without doing this review) than the death results,, but the death results could be used as an example.	The sentence has been shortened. It was determined that the caution is warranted to discourage people from inappropriately extrapolating, even if it should be obvious that they shouldn't.
Peer Reviewer #9	General	Well-structured but not sure it will impact policy or practice.	Thank you.