Comparative Effectiveness Research Review Disposition of Comments Report

Research Review Title: *Pulmonary Arterial Hypertension: Screening, Management, and Treatment*

Draft review available for public comment from August 31 to September 28, 2012.

Research Review Citation: McCrory DC, Coeytaux RR, Schmit KM, Kraft B, Kosinski AS, Mingo AM, Vann LM, Gilstrap DL, Hargett CW, Lugogo NL, Heidenfelder BL, Posey R, Irvine RJ, Wing L, Pendergast K, Dolor RJ. Pulmonary Arterial Hypertension: Screening, Management, and Treatment. Comparative Effectiveness Review No. 117. (Prepared by the Duke Evidence-based Practice Center under Contract No. 290-2007-10066-I.) AHRQ Publication No. 13-EHC087-EF. Rockville, MD: Agency for Healthcare Research and Quality. April 2013. 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=1414
Published Online: April 25, 2013

Commentator & Affiliation	Section	Comment	Response
Public Reviewer – American College of Chest Physicians (ACCP) 3	Abstract	*"No data are available regarding combined echocardiography and biomarker screening in asymptomatic patients at high risk for PAH. However, based on one good-quality prospective cohort study, biomarker testing with NT-proBNP may be useful in ruling out PAH in patients with symptoms suggestive of PAH who have elevated systolic pulmonary artery pressure (sPAP) by echocardiography."* "May be useful" is not very helpful. Second I would never rule out PAH in patients with symptoms suggestive of PAH who have elevated systolic PAP by echocardiography with a biomarker test NT-proBNP. I would perform pulmonary artery catheterization.	We agree that this study is affected by some potential bias in the way the cohort was assembled and the way the reference standard was applied. Furthermore, it is so small that numerical estimates are imprecise. However, it does remain one of the only studies that gave information on the joint distribution of echocardiography and BNP in diagnosing precapillary PH, and given the focus of our key question, it has to be featured rather prominently. Therefore, even though we retain the same sentence in the results portion of the structured abstract, we modified the conclusion section with a much more cautious conclusion as follows: "Further confirmation is needed to determine if echocardiography and the biomarker NT-proBNP in combination are sufficiently accurate to rule out PAH when testing symptomatic patients; in asymptomatic populations, more research is needed to draw conclusions regarding their effectiveness for screening" It seems fair to call for replication before widespread adoption of this strategy as there is a lot of uncertainty due to small numbers of subjects and confounding with ECG evidence or RVH. We have amended the discussion of echocardiographic testing, particularly about screening asymptomatic persons at high risk, to describe this issue. In the studies we evaluated, we attempted to estimate the proportion of patients who failed to have a measureable TR jet; however, this statistic was often unreported. Also, many studies used the presence of a measureable TR jet as an inclusion criterion and failed to detail the number of patients excluded for this reason.
Public Reviewer - ACCP 3	Abstract	*"Eighty studies involving 7224 patients evaluated biomarker tests, echocardiography, or both to evaluate severity or prognosis and followed progression of disease or response to therapy."* The statements do not incorporate recent derivation and validation of clinical prediction rules for survival (REVEAL Risk Calculator).	We now reference the Benza study in the Discussion section of the main report and in the KQ 2 section.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer - ACCP 3	Abstract	*"Although no treatments demonstrate a consistent mortality reduction, many are associated with improved 6MWD, and some are associated with reduced hospitalization rates."* No mention of time to clinical worsening (TTCW), only hospitalization. The review appears to miss combined endpoints such as TTCW which includes death and transplantation.	Composite endpoints are problematic to assess if individual endpoints making up the composite are not described. Composite endpoints, in essence, assign equal importance to different events in the composite. Furthermore, they are hard to compare when defined differently among studies. We assessed mortality and hospitalization separately; however, outcomes such as transplantation were even rarer than death and could not be examined separately. See Ferriera-Gonzalez I, Busse JW, Heels-Ansdell D, et al. Problems with use of composite end points in cardiovascular trials: systematic review of randomized controlled trials. BMJ 2007 doi:10.1136/bmj.39136.682083.AE (published 2 April 2007).
Public Reviewer - ACCP 10	Abstract	The statement that combination therapy resulted in reduced hospitalization is drawn from a very small # of patients with great heterogeneity; again any such statement would have to be muted in enthusiasm by this….	We have removed the statement about combination therapy and reduced hospitalization in the Abstract.
Peer Reviewer 2	Executive Summary	ES-3, lines 9-14 – I would urge caution re the statement on vasodilator response predicting responsiveness to CCBs. This is based on one retrospective study and has not been validated despite its wide acceptance in the field. Although this is not one of the questions being addressed, presenting the statement in the context of this document seems like an endorsement. Either soften the language by making it clear that this is a current practice not addressed by the current document, or just delete it.	We have reworded to state, "Calcium channel blockers are associated with long-term (>1 year) improvements in hemodynamics and functional status in most of those patients who show acute vasoreactivity testing response; however, acute vasoreactivity is seen in a minority of patients tested."
Peer Reviewer 2	Executive Summary	ES-14 and throughout document – it's Wood, not Woods units. Paul Wood, the British Physiologist after whom the unit is named, will not rest as peacefully if we get his name wrong.	Thank you; we have corrected this error.
Peer Reviewer 2	Executive Summary	ES-18, lines 56,57 – Delete "with" between monotherapy and PAH	We have corrected this typo.
Peer Reviewer 2	Executive Summary	Table E – Note morality outcome in ERAs v placebo and prostanoids v placebo boxes. Also, it's not endothelin agonists, which would not be desirable therapies, but rather Endothelin Receptor Antagonists (or ERAs).	We have corrected these errors.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 2	Executive Summary	ES-22, bottom, and ES-28, top, as well as P124 in the summary for document – The conclusion re adverse effects do not jibe with the Forest plots or with clinical experience. Drugs may not have differed in likelihood of causing diarrhea as per Fig 34, but there are no studies listing ERAs alone. Clinical experience indicates that ERAs are rarely associated with diarrhea and the lack of studies listed for ERAs alone should be acknowledged. Again, clinically, there is no way that PDESIs cause more peripheral edema that ERAs. Looking at Fig 35, I can't understand how the contrary conclusion that the authors draw is justifiable. The 3 ERA studies showed more edema in 3, and the 2 PDESI studies listed are combinations with PGI2 in one and an ERA in the other. Also, the association with prostanoids and cough is for inhaled prostanoids. Please re-examine Fig 38 with this in mind. Again, clinical experience indicates that there is no association between infusion prostanoids and cough. The problem here is that these studies are highly selected and reporting of side effects in them is not standardized and spotty at best. The authors need to temper their conclusions and acknowledge these limitations. The conclusions they're drawing are simply wrong and modification is mandatory.	We revised the text to better match the findings reported in the forest plots. Regarding the diarrhea adverse effect data, we now include in the forest plot the single study of an ERA that reported the incidence of diarrhea and its results were inconclusive. Regarding peripheral edema adverse effects data, we added the sentence, "This finding, however, does not necessarily indicate that phosphodiesterase inhibitors are associated with a higher incidence of peripheral edema relative to endothelin receptor antagonists or prostanoids." In addition, we moderated the conclusions as follows: "Phosphodiesterase-5 inhibitors and prostanoids were associated with flushing, while data on endothelin receptor antagonists were inconclusive. Phosphodiesterase-5 inhibitors had a significant association with peripheral edema while data on prostanoids and endothelin receptor antagonists were inconclusive." We revised the forest plots and associated text for all adverse events, including cough, to generate separate estimates for all prostanoids and inhaled prostanoids. We revised the discussion to include the limitation associated with the limited, unstandardized, and nonsystematic reporting of adverse events.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 2	Executive Summary	ES-25, and final conclusions – The recommendations re future research seem gratuitous unless the authors acknowledge some of the very real practical limitations related to some of the recommendations. Specifically, PAH is an orphan disease and recruitment is a major challenge. This limits the number of trials and endpoints that can be tested (ie survival). Because the prevalence of the disease in the general population is so low, screening of an asymptomatic population is senseless unless there are risk factors such as scleroderma, portal hypertension, catecholeminergic drug abuse or a family history to justify it. Longer-term trials are limited by patient numbers and ethical concerns that placebo control groups may lose ground that they will not make up unless they are some form of therapy. Head to head trials are limited by the need to have very large numbers to be powered adequately and the lack of interest or support on the part of the NIH (due to cost) or industry (due to the risk that their approved drug may not fare as well as a competitor's).	We added mention of specific high-risk populations (beyond systemic sclerosis) that may be candidates for screening test studies, including patients with HIV, sickle cell anemia or trait, portal hypertension, family history of PAH or catecholaminergic drug use. Given recent data that the principle surrogate outcome measure 6MWD explains little of the variance in long-term outcomes, it seems to us that either better surrogates need to be identified and developed or longer term outcome studies need to be used. While we agree that funding issues may preclude addressing some of the stated research priorities, we nevertheless believe that these are important research gaps that deserve mention.
Public Reviewer - ACCP 3	Executive Summary	*"The causes of PAH are numerous and are listed in Table A, taken from the Third World Symposium on PAH (2003)."* Out of date reference. Use Dana Point 2008.	We have updated our reference to Dana Point 2008, described in Simonneau G, Robbins IM, Beghetti M, et al. Updated clinical classification of pulmonary hypertension. J Am Coll Cardiol 2009; 54: S43-54.
TEP Member 5	Executive Summary	After reading the report. I think that it is prudently conducted and written. As a key informant and technical expert panelist, I think they took to heart our guidance and put those things into the methodology and conduct of the project. I had at the time suggested, but the majority of panelists disagreed, that the role of calcium channel blockers be included and I still believe that they should but that ship has sailed and I can appreciate that given the limited time and the limited funding to conduct the project, that this was not possible. I liked the clear guidance on BNP testing and the drug therapy sections. I think that for KQ1 key points that they could spell out the abbreviations like they do in other key points sections because if a reader wants the bottom line, they would be confused and have to go into the text. Also, with shading in the tables for insufficient versus low, medium, or high they may need to use a legend for 508 compliance. Finally, in their modified PRISMA I wonder if they shouldn't further break down where the studies were used qualitatively versus quantitatively.	Thank you. We have spelled out abbreviations at first use. Also, a list of abbreviations is provided at the end of the full report. Shading question to be addressed by AHRQ. Regarding the modified PRISMA: we ultimately used meta-analyses as a source for primary studies and as a basis for comparison with "what is already known" in the discussion section. We did not end up including any quantitative estimates from published meta-analyses in our analyses.

Commentator & Affiliation	Section	Comment	Response
TEP Member 7	Executive Summary	ES-1: The WHO classification of pulmonary hypertensive diseases referred to in the introduction was updated in 2008 and published in 2009 (Simonneau G, Robbins IM, Beghetti M, Channick RN, Delcroix M, Denton CP, Elliott CG, Gaine SP, Gladwin MT, Jing ZC, Krowka MJ, Langleben D, Nakanishi N Souza R. Updated clinical classification of pulmonary hypertension. J Am Coll Cardiol. 2009 Jun 30;54(1 Suppl):S43-54.	We have updated our reference to Dana Point 2008, described in Simonneau G, Robbins IM, Beghetti M, et al. Updated clinical classification of pulmonary hypertension. J Am Coll Cardiol 2009; 54: S43-54.
TEP Member 7	Executive Summary	ES-5: The two medications most recently approved by the U.S. Food and Drug Administration for PAH are inhaled treprostinil (July 2009) and tadalafil (May 2009) not ambrisentan (June 2007)	We appreciate this comment and have revised the Executive Summary.
Public Reviewer - ACCP 3	Executive Summary	*"Table A. Clinical classification of pulmonary hypertension."* Table A is out of date. Use Dana Point 2008.	We have updated our reference to Dana Point 2008.
Public Reviewer - ACCP 1	Executive Summary	ES-1 Would suggest referencing the Dana Point classification scheme in the Background, as opposed to the Venice system.	We have updated our reference to Dana Point 2008, described in Simonneau G, Robbins IM, Beghetti M, et al. Updated clinical classification of pulmonary hypertension. J Am Coll Cardiol 2009; 54: S43-54.
Public Reviewer - ACCP 1	Executive Summary	ES-3 Would recommend using the current definition of acute vasoreactivity.	We have reworded to state "Calcium channel blockers are associated with long-term (>1 year) improvements in hemodynamics and functional status in most of those patients who show acute vasoreactivity testing response; however, acute vasoreactivity is seen in a minority of patients tested."
Public Reviewer - ACCP 1	Executive Summary	In the last paragraph on page ES-3, a typo results in stating <6 mmHg, when this should read 36 mmHg.	We have corrected this in the final report.
Public Reviewer - ACCP 3	Executive Summary	*"The symptoms of PAH can be insidious and nonspecific and may include shortness of breath, fatigue, weakness, chest pain, loss of consciousness, and abdominal distention."* Symptoms also include syncope (usually exertional), leg swelling, and hoarseness (Ortner's syndrome). I would use syncope, since loss of consciousness occurs with hypoglycemia, drug overdose, etc.	We have changed *loss of consciousness* to *syncope* and add *leg swelling.*
Public Reviewer - ACCP 3	Executive Summary	*"RHC has traditionally been the means by which patients' clinical course is monitored; however, transthoracic echocardiography has emerged as a possible alternative."* There are many modalities for following PAH, not just pulmonary artery catheterization.	We chose to focus on echocardiography and biomarkers because there is particular uncertainty about the extent to which these could substitute for the reference standard of RHC in the monitoring process.
Public Reviewer - ACCP 3	Executive Summary	*"…although TAPSE and pericardial effusion have been proposed"* TAPSE has not been defined before the abbreviation is introduced	We have added the callout for TAPSE, tricuspid annular plane systolic excursion, at first use in the Executive Summary.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer - ACCP 3	Executive Summary	*"Calcium-channel blockers are associated with long-term (>1 year) improvements in hemodynamics and functional status in about half of the minority of patients who show acute vasoreactivity testing response."* This statement appears completely misleading because the definition of acute vasoreactivity in the study was a 20% decrease of PVR – an out of date definition. A much higher proportion of patients who meet the current definition of acute vasoreactivity (decrease of at least 10 mmHg to mean PA pressure less than 40 mmHg have a sustained response to treatment with high dose calcium channel blockers.	We have reworded to state "Calcium channel blockers are associated with long-term (>1 year) improvements in hemodynamics and functional status in most of those patients who show acute vasoreactivity testing response; however, acute vasoreactivity is seen in a minority of patients tested."
Public Reviewer - ACCP 3	Executive Summary	*"Figure B. Analytic framework for KQ 3:* This framework does not consider lung transplantation as an outcome.	We have added transplantation (lung or heart-lung) to the analytic framework as an outcome.
Public Reviewer - ACCP 3	Executive Summary	*"Table B. "NT-proBNP has variable sensitivity (range, 56% to 100%) for diagnosing PAH; uncertain performances for ruling in PAH. NT-proBNP has variable specificity (range, 24% to 95%); uncertain performance for ruling out PAH."* These statements lack clarity; a sensitive test allows me to rule out disease (few false negative tests) not rule in A specific test allows me to rule in disease (few false positive tests) not rule out.	We have corrected these statements in Table B.
Public Reviewer - ACCP 3	Executive Summary	*"presence of RV strain on ECG and serum NT-proBNP >80 pg/ml had a sensitivity of 100 percent and specificity of 19 percent for diagnosis of PAH based on RHC reference standard."* These statements appear incorrect; The sensitivity of this combination for PAH should be low. The specificity should be high.	We think the error was with the word "and" in that statement; the reported sensitivity and specificity is not for the combination of presence of RV strain AND serum NT-proBNP>80, but rather the presence of RV strain OR serum NT-proBNP>80. Conversely, the absence of RV strain on ECG AND the lack of elevated serum NT-proBNP (≤80 pg/mL) has 100% specificity and lower sensitivity. We have rephrased as follows: "… the presence of either RV strain on ECG or serum NT-proBNP >80 pg/mL had a sensitivity of 100 percent and specificity of 19 percent for diagnosis of PAH based on RHC reference standard."
Public Reviewer - ACCP 3	Executive Summary	*"BNP level is a poor predictor of mortality (high SOE)"* I would emphasize BNP "alone" is a poor predictor of mortality; to distinguish from BNP as part of a combined modality "risk score".	New data and a re-analysis that takes into account whether BNP levels were analyzed after logarithmic transformation suggest that log BNP level is a good predictor of mortality. Therefore, this key finding has been rephrased as follows: "Increase in level of log-transformed BNP is a strong predictor of mortality (moderate strength of evidence)."

Commentator & Affiliation	Section	Comment	Response
Public Reviewer - ACCP 3	Executive Summary	*"Table D. Summary of strength of evidence and effect estimates for the use of echocardiography or…"* Is there evidence related to TAPSE? Even if there is not, the table should include TAPSE as an echocardiographic measurement and state that there is no evidence.	We have included a meta-analysis of association between TAPSE and hazard ratio in the final report and added the results to Table D of the Executive Summary.
Public Reviewer - ACCP 3	Executive Summary	*"Table E. No clear differences in morality (few studies, fes deaths lead to wide CIs) No clear differences in morality (few studies, few deaths lead to wide confidence intervals"* An entertaining typo! Thank you.	We have corrected the typo.
Public Reviewer - ACCP 3	Executive Summary	*"Ínstead, we focus on available studies that evaluate the ability of echocardiography or biomarkers to assess the severity of PAH, to predict events such as lung transplantation or death"* How do you know that findings on echocardiogram or biomarker results did not influence decisions to refer for transplantation or enter hospice care, etc. i.e. influence the outcome of interest rather than predict the outcome of interest?	We agree that there may be many reasons for an association between test results and clinical outcomes; however, our goal was merely to describe the strength of association, not to flesh out the causal links involved. The design of included studies precludes that more ambitious goal. We have added text in the Discussion section (Limitations of the Evidence Base) regarding the possibility of this type of bias. "Assessing the prognostic value of biomarkers or echocardiographic parameters for such outcomes as need for transplantation may be biased since all these studies were observational, lacked blinding, and the predictors may have influenced clinical decisions about management or referral for transplantation."
Public Reviewer - ACCP 3	Executive Summary	*Ín patients treated for PAH with prostanoids, endothelin antagonists, or phosphodiesterase inhibitors, current evidence is inconclusive regarding a reduction in mortality with either monotherapy or combination therapy (insufficient SOE)."* This statement appears to be at odds with clinical experience. For example, I do not believe you could find anyone who would believe it ethical to randomly assign idiopathic PAH patients to intravenous epoprostenol or placebo and test the hypothesis that there is no difference in survival.	Our review is an assessment of existing evidence and neither a prescription for clinical practice nor a call for trials to address this question. The available evidence allows us to test the hypothesis that active treatment is associated with improved outcomes (including mortality) compared with placebo or standard treatment, despite the fact that trials were not designed specifically to test for a mortality benefit.
Public Reviewer - ACCP 4	Executive Summary	ES-3, typo- last paragraph RVSP<6 should be <36	We have corrected this in the final report.
Public Reviewer - ACCP 4	Executive Summary	ES-5 first paragraph, "The 2 medications most recently approved… are inhaled treprostinil and ambrisentan." This should be modified to change ambrisentan for tadalafil	We have modified the text.
Public Reviewer - ACCP 4	Executive Summary	Page 1, Table 1. Clinical Classification is from 2004, not 2009	We have updated our reference to Dana Point 2008.
Public Reviewer - ACCP 5	Executive Summary	ES-3 has a typo 4th line from the bottom (should be 36 rather than 6 mmHg)	We have corrected this text.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer - ACCP 9	Executive Summary	ES-12: important to know what you used as a "criterion standard diagnosis"	Our inclusion criteria require that studies confirm PAH diagnosis with RHC at least for a majority of test-positive patients. We now use the phrase "reference standard diagnosis based on RHC."
Public Reviewer - ACCP 9	Executive Summary	The lack of references within the text through most of the executive summary markedly diminishes its value.	We have adhered to the AHRQ guidelines for use of references in the Executive Summary. The reader is referred to the full report for references.
Public Reviewer - ACCP 9	Executive Summary	ES-23 "…or with no NHYA class II or IV symptoms…" – this is not clear, and I suspect a typo as such a grouping makes no sense.	We have corrected this error as follows: "with no NYHA class III or IV symptoms …"
Public Reviewer - ACCP 10	Executive Summary	There may be too much stock being placed in the single study by Bonderman, et al in stating that "based on one good-quality prospective cohort study that the biomarker NT-proBNP may useful in ruling out PAH in patients with symptoms suggestive of PAH and elevated sPAP on an echo. While the authors later admit strength of evidence is low, it does not seem to prevent them from recommending this approach nonetheless. I would not endorse such a sweeping recommendation after reading that study.	We agree that this study is affected by some potential bias in the way the cohort was assembled and the way the reference standard was applied. Furthermore, it is so small that numerical estimates are imprecise. However, it does remain one of the only studies that gave information on the joint distribution of echocardiography and BNP in diagnosing precapillary PH, and given the focus of our key question, it has to be featured rather prominently. It seems fair to call for replication before widespread adoption of this strategy as there is a lot of uncertainty due to small numbers of subjects and confounding with ECG evidence or RVH.
Public Reviewer - ACCP 10	Executive Summary	The conclusion that BNP alone could not serve as an accurate surrogate marker for disease really does not surprise me and if this Key Question is used, I suppose pointing this out to caregivers would be reasonable.	Thank you for your comment.
Public Reviewer - ACCP 10	Executive Summary	Pericardial effusion as the strongest predictor of mortality also not new news but if this question is to be use this can be reinforced with strength of evidence as outlined.	Thank you for your comment.
Public Reviewer - ACCP 10	Executive Summary	The statement that PDE-5 inhibitors and ERAs use are associated with lower rates of hospitalization while a lack of such association was found with prostacyclin use is misleading and may in fact reflect severity of disease as caregivers are likely to use former agents in less sick individuals who by definition likely have lower risk of need for hospitalization in the first place. I did not see that an effort had been made to take this into account in the analysis (ie RA pressures, CI's, functional class status, etc). As such any such statement needs to be guarded with recognition that we may not be comparing apples to apples here.	We agree with the reviewer that the referenced statement was misleading. In the revised report we characterize these data as follows: Key Point: "Endothelin antagonists and phosphodiesterase inhibitors were associated with lower incidence of hospitalization when compared with standard therapy or placebo (moderate strength of evidence). Current evidence is inconclusive regarding a reduction in hospitalization associated with treatment with prostanoids (insufficient strength of evidence)."

Commentator & Affiliation	Section	Comment	Response
Public Reviewer - ACCP 10	Executive Summary	Again the conclusion statement that "echocardiography and the biomarker NT-proBNP in combination may be sufficiently accurate to rule out PAH when screening symptomatic patient is too strong a conclusion and needs to be tuned down. Second statement in conclusion that "more research in screening asymptomatic patients" is also obvious to those in the field.	We agree that this study is affected by some potential bias in the way the cohort was assembled and the way the reference standard was applied. Furthermore, it is so small that numerical estimates are imprecise. However, it does remain one of the only studies that gave information on the joint distribution of echocardiography and BNP in diagnosing precapillary PH, and given the focus of our key question, it has to be featured rather prominently. It seems fair to call for replication before widespread adoption of this strategy as there is a lot of uncertainty due to small numbers of subjects and confounding with ECG evidence or RVH.
Public Reviewer - ACCP 10	Executive Summary	I think any statements made on the impact of combination therapy on mortality are likely to be vapid and inane as none of the small 3 of studies were designed or powered to assess this outcome. The attempt to address this indirectly by combining these studies is very dangerous ground as the patient populations studied were significantly different and the short time of follow-up makes any such attempt misguided I think.	We acknowledge that there are important differences between the combination therapy versus monotherapy studies. However, the Key Question involves either monotherapy or combination therapy. Furthermore, mortality is the most important outcome, and the small size of the studies is not a valid reason to skip the mortality analysis. The net result was inconclusive, so we drew no conclusions. We agree that there are important caveats to combining trials using similar strategies but different drugs; however, under these circumstances, it seems that a test of the null hypothesis of no effect in a meta-analysis is reasonable. We have added a description of our approach to meta-analysis at the beginning of the KQ 3 Results section to explain this decision and frame such an analysis with the appropriate caveats. This text notes that studies "randomized patients who had previously received monotherapy to either continued monotherapy with that drug or continued therapy with that drug plus the addition of a second drug. For the purpose of this report, we consider these studies to represent a comparison of combination therapy with monotherapy, with the understanding that this study design does not address the question of whether initiating two drugs is superior to initiating a single drug to treatment-naïve patients."

Commentator & Affiliation	Section	Comment	Response
Public Reviewer - ACCP 10	Executive Summary	Regarding the conclusion that combination therapy resulted in greater 6 MWD than monotherapy there is no recognition on part of authors as to what constitutes a clinically meaningful difference in 6 MWD and any statement from our committee would have to tackle this question (perhaps in part from guidance from the recent paper by Mathai S, et al?)	We have added a supporting citation for an emerging consensus regarding MID for 6MWT of 33 meters (Mathai et al 2012).
Public Reviewer - ACCP 10	Executive Summary	Again on page 18 the statement that PDE-5 and ERAs use were associated with placebo or conventional therapy is probably ok, but the assertion that prostanoid use was not is again confounded by the severity of illness of the later patients.	We cited two studies of prostacyclin's effect on hospitalization. While it is possible that the RCTs that evaluated prostacyclins recruited sicker patients compared with studies of PDE-5 inhibitors or ERAs, the confidence intervals around the OR for hospitalization associated with prostacyclins were wide and included 1. This could also be accounted for by type two error. We agree with the reviewer that the referenced statement was misleading. In the revised report we characterize these data as follows: Key point "Endothelin antagonists and phosphodiesterase inhibitors were associated with lower incidence of hospitalization when compared with standard therapy or placebo (moderate strength of evidence). Current evidence is inconclusive regarding a reduction in hospitalization associated with treatment with prostanoids (insufficient strength of evidence)."
Public Reviewer - ACCP 10	Executive Summary	The statement that "cough is associated with prostanoid use" must be clarified to "inhaled prostanoid use" as this is not a finding in trials with infused prostonoids; the report does not make this distinction and should, Likewise the statement that jaw pain is associated with prostanoid use should be "infused prostanoid use" since inhaled prostanoids are not really associated with this side effect.	We have revised the forest plots for all of the adverse events analyses to generate separate estimates for all prostanoids vs. inhaled prostanoids only, and we have revised the text accordingly.
Public Reviewer - ACCP 10	Executive Summary	The remainder of the conclusions seem reasonable, but again the clinical importance of much of these will need to be discussed by the panel and the frequent statement of "heterogeneity of data" with "insufficient strength of evidence" seems to be the common theme here.	Thank you for your comment.
Peer Reviewer 1	Introduction	Excellent overview and review of the background of PAH.	Thank you.
Peer Reviewer 1	Introduction	Page 12 line 33 should read "a transthoracic" rather than "an transthoracic."	We have corrected this error.
Peer Reviewer 1	Introduction	Page 12 line 52 should read "RVSP<36" rather than "RVSP<6."	We have corrected this error.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 1	Introduction	Page 22 line 8: suggest spelling out the abbreviation "SOE" strength of evidence at the legend of the table.	We have spelled out the abbreviation in headings and captions as well as including it in the footnotes to the tables where SOE is used.
Peer Reviewer 3	Introduction	The key questions were explicitly stated.	Thank you for your comment.
TEP Member 1	Introduction	Concise overview of the condition, limitations of analysis in this orphan disease, and rationale for this work.	Thank you for your comment.
TEP Member 2	Introduction	Generally appropriate	Thank you for your comment.
TEP Member 3	Introduction	The introduction is concise and easy to understand	Thank you.
TEP Member 4	Introduction	Background information is sufficient. Inclusion of the clinical classification table is important, but this is NOT the most up-to-date clinical classification schema. The Dana Point classification should be included instead, as there are differences between the 2003 and 2008 classifications. For reference: Simonneau G, Robbins IM, Beghetti M, et al. Updated clinical classification of pulmonary hypertension. J Am Coll Cardiol 2009; 54:543-54.	We have updated our reference to Dana Point 2008, described in Simonneau G, Robbins IM, Beghetti M, et al. Updated clinical classification of pulmonary hypertension. J Am Coll Cardiol 2009; 54: S43-54.
TEP Member 4	Introduction	Also, the citation to reference #3 for the clinical classification origin is not correct. It would be reference #4 in the bibliography.	We have updated our reference to Dana Point 2008.
TEP Member 5	Introduction	Excellent sumary in my opinion. Brings people up to speed in an understandable formatfor various stakeholders. Would have been easy to lose people in the minutia	Thank you.
TEP Member 6	Introduction	Given the extensive use of echo in this disease state, the summary does not contain key information. For example, PASP as a reliable marker of disease severity or prognosis is dependent upon RV systolic function and stroke volume. This is a key issue since pulmonary htn by echo or by cath can underestimate severity of PVR when RV function is poor. There probably needs to be some mention of more recent state-of-the-art echo based methods for evaluating RV systolic function, particularly with regard to which are load-dependent and load-independent. Biomarkers - most can be substantially influenced by other comorbidities such as hypertensive heart disease or ischemic heart disease. Treatment strategies are geared toward chronic treatment but misses some acute therapies.	We added to "Role of Echocardiography" the following text in both Executive Summary and introduction: "Furthermore, sPAP is dependent on right ventricle (RV) systolic function and stroke volume. In later stages of PH, RV function deteriorates, which can lessen the degree of sPAP elevation and lead to an underestimate of pulmonary vascular resistance (PVR). More recent echocardiographic-based methods have focused on evaluating RV systolic function."
TEP Member 8	Introduction	Background: excellent list of etiologies of PHTN.	Thank you for your comment.
TEP Member 8	Introduction	Page 11 line 56: I prefer the term reference standard; "gold standard" is a poor term. Later on in the report reference standard is used.	We changed from "gold standard" to "reference standard" throughout the report

Commentator & Affiliation	Section	Comment	Response
TEP Member 8	Introduction	Line 57: … an invasive but generally safe… is how I would phrase it Page 12… because of its ubiquitous availability, cost effectiveness, complete safety it highly desireable to encourage further development.	We have changed our language in this section to reflect these comments.
TEP Member 8	Introduction	P22 No mention of mean pressure from echo. There are at least three ways to do this and the literature needs to be expanded here. When you consider that your are comparing peak systolic pulmonary pressure from echo with RHC and that RHC uses peak pressure as the definition of the dx of PHTN (>25mmHg), exploration of echo derived mean pressure would seem to hold great promise. That has certainly been our experience. Also, in our experience, we find the exercise response of pulmonary pressure to be a promising approach to use with echocardiography but there is little if any data. I would mention this as a promising approach.	We did evaluate accuracy of echo methods for estimating mPAP. However, there were too few data to support a SOE rating for this parameter in the SOE table, hence its omission. We have added points to the research gaps section: - Further development of data on the use of echo to measure exercise response to sPAP - Further development of echocardiographic estimation of mPAP, which would better align with the diagnostic criteria for PAH
Public Reviewer - ACCP 6	Introduction	The introduction is concise and easy to understand	We appreciate your comment.
Peer Reviewer 1	Methods	The authors provide a clear and thorough description of their methods to review the literature.	Thank you for your comment.
Peer Reviewer 1	Methods	Key Question 2, evaluating management and outcomes in PAH, includes hemodynamic parameters as an intermediate outcome measure. Hemodynamic parameters at right heart catheterization are measurements similar to echocardiographic hemodynamic measurements – these can be used in management and prognosis but are not outcomes in themselves. As an analogy, blood pressure measured noninvasively by a BP cuff does not predict an outcome of BP measured through an arterial line. They are both tests that measure the same thing. Either measurement can be used to predict short term outcomes such as symptoms, walk distance, hospitalizations, or other clinical events.	We agree that the use of hemodynamic measurements as surrogate outcomes is problematic; nevertheless, as the reviewer indicates, there are data to support their prognostic values with respect to clinically important outcomes. One of the main aims of this report was to determine if there is strong enough correlation with these surrogate outcomes that would support replacing invasive with noninvasive procedures in the management of PAH. We therefore sought data on the accuracy of noninvasive estimates of hemodynamics as well as data on the association with mortality and other health outcomes.
Peer Reviewer 1	Methods	While the authors gathered many articles comparing the diagnostic accuracy of lab tests or echocardiography versus cardiac catheterization, cath is not an outcome measure. The description of these studies is valuable, but might rather be considered as the diagnostic ability of noninvasive versus invasive parameters among those being followed for PAH.	One of the main aims of this report was to determine if there is strong enough correlation with these surrogate outcomes that would support replacing invasive with noninvasive procedures in the management of PAH. We therefore sought data on the accuracy of noninvasive estimates of hemodynamics as well as data on the association with mortality and other health outcomes.
Peer Reviewer 2	Methods	Methods are sound.	Thank you for your comment.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 3	Methods	The inclusion and exclusion criteria are clearly defined and are resonable, and the diagnostic criteria are appropriate.	Thank you for your comment.
TEP Member 1	Methods	Comprehensive and appropriate methodological approach. I saw no glaring omission of a relevant study.	Thank you for your comment.
TEP Member 2	Methods	Very good	Thank you for your comment.
TEP Member 3	Methods	The methods employed are characteristic of systematic reviews and included data abstracted from all published texts relevant to the subject matter. The data was rated for quality and applicability and graded for strength of evidence. Random effects models were used to compute summary estimates of the effect were similar studies provided these estimates.	Thank you for your comment.
TEP Member 4	Methods	Exclusion and inclusion criteria are justifiable. No issues with the methodology, but I defer judgment to other Reviewers who may have more expertise in bioinformatics approaches.	Thank you for your comment.
TEP Member 5	Methods	Prudent methods.	Thank you for your comment.
TEP Member 6	Methods	Very clear. Was there any effort made to exclude duplicate studies of the same population where there were slight variations in the Aims or Methods?	We did attempt to identify and exclude duplicate publications. When identified we used the most complete or most recent as judged for our purposes on a case by case basis. Sometimes identifying such duplicates is difficult.
TEP Member 7	Methods	The inclusion criteria, search strategies, outcome measures and statistical methods are appropriate.	Thank you for your comment.
TEP Member 8	Methods	No problem with these factors.	Thank you for your comment.
Public Reviewer - ACCP 6	Methods	The methods employed are characteristic of systematic reviews and included data abstracted from all published texts relevant to the subject matter. The data was rated for quality and applicability and graded for strength of evidence. Random effects models were used to compute summary estimates of the effect were similar studies provided those estimates.	Thank you for your comment.
Peer Reviewer 1	Results	The authors provide a thorough literature search and the table formats allow overview of the articles in an organized manner. While it is difficult to accurately summarize an entire article in a few lines, the authors seem to have made a valuable attempt to do so. I believe the data is presented clearly.	Thank you for your comment.
Peer Reviewer 2	Results	Figures A and B, and 1 and 2 – the distinction between intermediate and long term outcomes is artificial. Would eliminate the categorization. All are of interest in both contexts.	We agree that this separation is problematic, with the exceptions of hemodynamic parameters, which are clearly a surrogate outcome, and mortality, which is clearly an ultimate health outcome; hence our decision to keep them all in the same box. We have revised the analytic framework by eliminating the distinction between intermediate-term and long-term outcomes.
Peer Reviewer 2	Results	Main text, page 4 – lines 30 to 33 and 35 to 38 are redundant.	We have deleted the redundant text.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 2	Results	Page 38, Table 10 – define RIMP on table legend	We have added "RIMP=right ventricular index of myocardial performance" to the table abbreviations.
Peer Reviewer 2	Results	Many of the Forest plots (Fig 9 through Fig 13 for example) are difficult to read. Especially the numbers on the abscissa. These are not of publication quality.	We have produced higher resolution graphics for the final report.
Peer Reviewer 2	Results	Page 86, lines 46,47 – The term "secondary to" is no longer used for CTD-PAH. The term, according to Dana Point, is "associated with".	Consistent with our use of the 2008 Dana Point criteria, we have replaced "secondary to" with "associated with" when referencing suspected etiology of PAH.
Peer Reviewer 2	Results	Page 86, lines 49 to 51 – This is impossible to follow. CTEPH is a form of PH, it doesn't have secondary PAH. CTEPH is mentioned twice, suggesting 2 studies are being cited, but not clear in the sentence.	We have clarified the text as follows: "Two studies enrolled a minority of patients with PH other than PAH: one included patients with chronic thromboembolic PH (28%), and another included patients with PH owing either to lung disease or to chronic thromboembolic PH (37%)."
Peer Reviewer 2	Results	Page 106, Figures 29-31 – Why only 4 studies for mPAH and 5 for PVR and CI. It's hard to fathom how mPAH wasn't available from the Barst study if the other 2 were.	While many studies reported means, relatively few studies reported means and standard deviations or CIs required for calculation of effect sizes. For the revised report, we computed CIs utilizing SE and the 0.975-quantile from a t-distribution with degrees of freedom based on sample sizes of the two groups as described in the revised Methods. This made it possible to include more studies in the meta-analyses.
Peer Reviewer 2	Results	Forest plots for therapy – what is the significance of the Prostanoid row at the borrom of these plots. No author or sample sizes are listed and it would seem that these represent a sum signal. Please clarify.	We have revised all of the forest plots (using a different meta-analysis software program) to clearly indicate the summary estimates in each plot.
Peer Reviewer 3	Results	The results are clearly described in appropraite detail.	Thank you for your comment.
TEP Member 1	Results	Quite complete and readable. Mode of presentations of data (tables, figure, graphs and plots) are readable and easy to follow, and address the key questions to the extent the data allow.	Thank you for your comment.

Commentator & Affiliation	Section	Comment	Response
TEP Member 2	Results	Yes, very detailed in general. A couple of minor points. Cough is pointed out as a side effect of prostanoids. But this is specific to inhaled prostanoids and this should be clarified. I think it is a misrepresentation that ERA's and PDE's improve clinical worsening but prostanoids do not. The time to clinical worsening endpoint was first introduced in the BREATHE-1 trial. It was not an endpoint in the initial epoprosentol and treprostinil trials. In fact, had it been included in the Barst 1996 NEJM trial, there would be no doubt that it would hit that endpoint based on the mortality benefit alone. The way this is presented is deceiving to the reader who may not know these details.	We have revised the forest plots to generate separate estimates for all prostanoids versus inhaled prostanoids only, and we have revised the text accordingly. Very few eligible studies reported time to clinical worsening. We did not, therefore, analyze and report that outcome. Some, but not all, of the studies incorporated clinical worsening in their hospitalization outcome, but it appears that many studies reported hospitalization for any reason.
TEP Member 3	Results	Only one prospective study supported the use of biomarker testing with NT pro-BNP combined with echocardiography as possibly helpful in ruling out pulmonary hypertension in symptomatic patients. I believe the reason that there is only one prospective study, is that clinical experience has already proven that this strategy is not a helpful in the clinical arena. BNP levels alone have not been shown to be an accurate surrogate marker for disease severity even in CHF. More importantly, is the echo Doppler cannot be used as a screening tool in asymptomatic or minimally symptomatic individuals. The use of Doppler requires a well-defined tricuspid regurgitation jet which will occur only in patients with elevated pulmonary artery pressure and right ventricular dilation, (i.e. patients with relatively advanced disease). Asymptomatic and minimally symptomatic patients often do not have significant tricuspid regurgitation. They also found that the presence of a pericardial effusion was associated with increased mortality, but this observation has been made previously, and primarily refers to patients with scleroderma.	We have amended the discussion of echocardiographic testing, particularly about screening asymptomatic persons at high risk, to describe this issue. In the studies we evaluated, we attempted to estimate the proportion of patients who failed to have a measureable TR jet; however, this statistic was often unreported. Also, many studies used the presence of a measureable TR jet as an inclusion criterion and failed to detail the number of patients excluded for this reason.

Commentator & Affiliation	Section	Comment	Response
TEP Member 3	Results	The assessment of drug efficacy presents another constellation of challenges. Virtually all of the randomized clinical trials conducted for this disease were very short duration, 3 to 4 months, and utilized 6 min. walk as the primary endpoint. Because of the lax regulatory requirements for drug approval in PH, most did not include hemodynamic outcome data, or data on mortality. Head-to-head trials have not been conducted, and thus there is little possibility of being able to make judgments on relative efficacy. Thus it is no surprise that their study of comparisons between agents were inconclusive. Combination studies have not been conducted in a structured way since it would require the cooperation and support from two or more pharaceutical companies.	We agree with the reviewer's assessment.
TEP Member 4	Results	The amount of detail presented in the results seems appropriate. The characteristics of the studies are clearly described. The key messages are explicit and applicable. The figures, tables and appendices are adequate.	Thank you for your comment.
TEP Member 4	Results	Note: line 56-57 contains either a grammatical error or a word is missing: "Monotherapy with for PAH is associated with statistically…"	We have corrected the error.
TEP Member 5	Results	Reasonable results. Further emphasizing why some drugs in the same class could not be combined (like for the mortality endpoint) could be helpful to the reader. Perhaps a symbol in the table leading to a notation.	We have added a forest plot and accompanying analysis that combines drugs within class. The revised report combines studies of different drugs within the same class, with an accompanying forest plot for each outcome.

Commentator & Affiliation	Section	Comment	Response
TEP Member 6	Results	The detailed synthesis for KQ1 misses two key points that make this study less useful. First, echocardiographic experts would not call PASP >36 mm Hg if there is evidence for increased stroke volume which occurs commonly in patients with anxiety, liver disease, anemia, etc. Second, there is convention in echocardiography to estimate left sided filling pressures - this is meant to differentiate between pre and post capillary PH and should have been part of the study. The wide range of specificity for BNP in PH is entirely attributable to different inclusion criteria. Its role becomes limited the more the general population is included. FAC is a surrogate for RVEF - this is a recognized poor marker for RV function. There have been more recent studies on strain, TDI S', and TDI acceleration. It may or may not be of interest, but there are recent studies showing safety of contrast-enhanced ultrasound (with Definity) in pulmonary htn.	We added text to the Background to acknowledge the point about stroke volume vis a vis threshold for sPAP. We did not find studies that reported echocardiographic estimates of left-sided filling pressure to distinguish between precapillary and postcapillary PAH, although we sought studies that used echocardiography to make this distinction using combinations of echo parameter that might separately assess PAP, RAP, PCWP estimates, and RV function. This was identified as an area of future research. FAC is recommended in the ASE guidelines. TDI S' was omitted from the draft due to problems with the reliability of our data-gathering forms. We have corrected this error in the final report. We appreciate the reviewer bringing to our attention the recent article: Wever-Pinzon O, Suma V, Ahuja A, et al. Safety of echocardiographic contrast in hospitalized patients with pulmonary hypertension: a multi-center study. Eur Heart J Cardiovasc Imaging. 2012 Oct;13(10):857-62. PMID: 22427401. We did not identify any studies in PH looking at the perflutren lipid microsphere for either safety or diagnostic accuracy. In October 2007, the FDA placed a black box warning on the label of the perflutren-based agents Definity and Optison, contraindicating their use in patients with pulmonary hypertension. We were unaware that this was revised in 2011, however, the ban did apparently result in unavailability of data on the effect of Definity on improving the diagnostic accuracy of echo, hence, isolated data on its safety is of little use.
TEP Member 7	Results	The section is very detailed and clearly described. Most of the figures are helpful. The tables are cumbersome because of their size and are difficult to compare. It would be helpful to leave out some of the detail here or to provide smaller summary tables that could be more easily compared.	The Results section of the report must meet the needs of a wide audience from various backgrounds, and we have included a level of detail in the tables to address this. We hope that the more concise results in the Executive Summary are helpful in this case.

Commentator & Affiliation	Section	Comment	Response
TEP Member 7	Results	The key messages are explicit, but the strength of evidence is too low to make most messages applicable. In particular the following conclusions may be problematic: 1) For patients suspected of having PAH with elevated systolic pulmonary artery pressure (sPAP) by echo, additional testing with the biomarker N-terminal pro-B-type natriuretic peptide (NT-proBNP) may identify patients who do not have PAH compared with echo sPAP alone (based on one good-quality prospective cohort study) (low SOE). Although the limitation of only 1 study is mentioned and SOE is rated low, the overall message relayed is that a normal NT-BNP may obviate the need for right heart catheterization in patients with elevated PA pressure on echo. This may be construed as a recommendation. My conclusion from this report is that there is insufficient data to determine if NT-BNP or any other biomarker can obviate the need for further diagnostic evaluation. Current clinical guidelines suggest that BNP be used as only one of several factors to determine if right heart catheterization is necessary. Page 23. Predicting Incidence of PAH The 2 studies cited had low numbers of patients and both evaluated patients with SSc, making it difficult to extrapolate the high specificity to other WHO group 1 PAH patients. Page 23 Diagnosis of Prevalent PAH I disagree with the statement that “Differences between sensitivity and specificity estimates among these studies likely stem from the inclusion criteria in the study by Bonderman et al.25 in which all patients had elevated sPAP (>40 mmHg) by echocardiography, leading to a population with a high proportion of patients who had elevated NT-proBNP levels”. A better explanation would be the unusually high mPAP > 35 that was used for diagnostic reference criteria in this study that likely had the effect of sharply reducing the specificity.	While we agree that a high mPAP threshold for diagnosis of PAH in this study could affect the specificity as you describe, the high diagnostic threshold for mPAP noted in the table for Bonderman et al. is an error. The study used the standard threshold of mPAP>25mmHg.

Commentator & Affiliation	Section	Comment	Response
TEP Member 7	Results	2) For patients suspected of PAH, echocardiographic estimation of right ventricular systolic pressure (RVSP) (or tricuspid gradient [TG]) by tricuspid regurgitation jet velocity (TRV), sPAP by TRV and right atrial pressure (RAP), and pulmonary vascular resistance (PVR) by TRV/velocity-time integral right ventricular outflow tract (VTIRVOT) show reasonably good accuracy compared with right heart catheterization (RHC) (moderate SOE). This is seen again on page 54: " Systolic pulmonary artery pressure (sPAP) estimated by echocardiography shows good correlation with sPAP from RHC (moderate SOE)". Although I agree that there is considerable data to demonstrate that echo estimates of PA pressure correlate well with pulmonary artery catheterization, the phrase "show reasonably good accuracy compared with right heart catheterization" is simply too misleading and ignores the more important clinical point that in an unacceptably high percentage of cases, echocardiography is inaccurate and cannot be used to make the diagnosis or determine treatment response. This limitation is noted on page 42: "With a standard deviation of this magnitude, one would expect about 80 percent of echocardiography sPAP reading to fall within 10 mmHg of RHC sPAP; however, the large Reveal registry9 found that only 39.8 percent of echocardiography sPAP estimates were within 10mmHg of same-day RHC-measured sPAP, corresponding to a standard deviation of around 20 mmHg. The inability of echo estimates of sPAP to be within 10 mmHg for 20% of patients argues against the statement of "reasonably good accuracy". This is particularly true because the mean reduction in sPAP in response to therapy is only about 10 mmHg	We agree that the REVEAL registry findings are troublesome; however, this registry reported only patients PAH with high PAP; furthermore, the measurement error was correlated with the magnitude of PAP. Unfortunately, among patients with lower PAP, lower reliability of the presence of a TR jet may limit accuracy by a different mechanism. The inconsistency stems partly from concerns about the applicability of the REVEAL registry study to a different population. We expanded the discussion of these concerns. We also identified three other studies that report misclassification rates in a similar manner to the REVEAL study and tabulated and discussed these data (Table 12). Interestingly, the additional data indicated somewhat lower standard deviations than the REVEAL study, suggesting that the misclassification problem might be slightly smaller than REVEAL suggests.

Commentator & Affiliation	Section	Comment	Response
TEP Member 7	Results	3) Evidence comparing monotherapy with combination therapy for PAH permitted a conclusion for only one outcome: improved 6MWD with combination therapy compared with monotherapy (low SOE). Data supporting this conclusion come from studies in which patients on one PAH therapy are randomized to receive an additional PAH specific therapy or placebo. I disagree with the statement that following phrase on page 98 "…a trial of tadalafil with bosentan background therapy can demonstrate the efficacy of combination versus monotherapy". This stepwise addition of therapy does not test the efficacy of combination versus monotherapy. There is a strong selection bias for patients to enter this type of study when they are not responding to their background therapy. Thus, the improved results may simply be the result of the new therapy versus placebo and not the benefit of combined therapy versus monotherapy This point should be emphasized throughout K3. Presently there are simply no studies that adequately address the efficacy of combination therapy versus monotherapy	We appreciate and agree with the observation about monotherapy versus combination therapy. We have revised the report accordingly in several places, including the Abstract, Key Points, the Results for KQ 3, and the Conclusions. The revisions include the following Key Point: "In patients who have been receiving monotherapy, combination therapy appears to be moderately more effective than continuation of monotherapy for improving 6-minute walk distance (6MWD), with a magnitude of effect that is approximately equal to the estimated minimal important difference (MID) of 6MWD for PAH of 33 meters."
TEP Member 8	Results	Detail is excellent. There are a few papers from our group using exercise pulmonary pressure response in connective tissue disease that deserve mention. These study small numbers of subjects and need to be expanded.	We reviewed papers on this topic from the reviewer, but determined that these studies did not meet inclusion criteria because they either antedated the inclusion year or PAH diagnoses or lacked RHC verification.
Public Reviewer - ACCP 3	Results	*"We identified only one study (good quality) that gave data on the use of echocardiography and biomarkers in screening patients suspected of having PAH; Bonderman D, Wexberg P, Martischnig AM, et al. A noninvasive algorithm to exclude pre-capillary pulmonary hypertension. Eur Respr J 2011;37(5):1096-103. PMID: 20693249"* I do not have confidence in this study. First, what was the prevalence of disease is the prospective study population, and do they match the population that I evaluate? The study is from a single center in Europe.	We agree that this study is affected by some potential bias in the way the cohort was assembled and the way the reference standard was applied. Furthermore, it is so small that numerical estimates are imprecise. However, it does remain one of the only studies that gave information on the joint distribution of echocardiography and BNP in diagnosing precapillary PH, and given the focus of our key question, it has to be featured rather prominently. It seems fair to call for replication before widespread adoption of this strategy as there is a lot of uncertainty due to small numbers of subjects and confounding with ECG evidence or RVH.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer - ACCP 3	Results	*"In Table 5, reference 25; mPAP>35 mmHg PCWP<15 mmHg"* Is this an error? Should the criteria be mean PA pressure >25 mmHG? I think the mPAP is actually the systolic PA pressure estimated from echocardiography. Also the reference mean PA pressure threshold is not provided IN THE TABLE for references #58 AND #68	On review of Bonderman (2011), we agree that the RHC threshold for diagnosis is actually mean PA pressure >25 mmHg and PCWP <15 mmHg. Simeoni (2008) used a case-control design and did not describe the specific diagnostic criteria used. We updated the table to indicate NR to describe when no specific criteria were reported and noted the case-control design. Williams (2006) did not explicitly state the diagnostic criteria used at RHC, but referred to a previously published protocol (Mukerjee et al., 2003) which lists mean PAP >25mmHg at rest or exercise induced mPAP>30mmHg with PCWP <15 mmHg. We have revised the table to include these thresholds.
Public Reviewer - ACCP 3	Results	*"At baseline patients were either without any signs or symptoms suggesting PAH34 or with no NYHA class II or IV symptoms and echocardiographic estimate of sPAP less than 40 mmHg."* The part "with no MYHA class II or IV symptoms" does not make sense to me. Do you mean III or IV?	Indeed, the study by Allanore et al. excluded patients with NYHA class III or IV symptoms. We have corrected this error as follows: "with no NYHA class III or IV symptoms …"
Public Reviewer - ACCP 3	Results	*"Table 6. Biomarker levels by diagnostic group"* I do not have confidence in Table 6; I reviewed only the Bonderman article. The diagnostic criteria was MEAN PA pressure >25 mmHg at rest (NOT >35 mmHg).	We have corrected the text in Table 6.
Public Reviewer - ACCP 3	Results	*"Table 8. Diagnostic accuracy of echocardiographic parameters for diagnosis of PAH"* There is no diagnostic standard.	We are not asserting that echocardiographic parameters are the diagnostic standard; rather, our analysis seeks to assess the performance of echocardiographic parameters against the reference standard (i.e., right heart catheterization).
Public Reviewer - ACCP 9	Results	Table 5 (and elsewhere as needed): Among the the patients "referred for evaluation of suspected PAH" as noted for the Bonderman study, the validation group (121) and many of the others I suspect, were referred after an echo was performed, and presumably suggested PH. This is an important issue in evaluating to whom the results reported may be applied. I have not point checked all the other references used in this regard, but would hope the authors would make sure of this issue as again, it related to what may and may not be concluded about the generalizability of the findings.	Some of the studies are certainly affected by verification bias, incorporation bias, referral filter bias, or combinations of these. These biases can be expected to inflate the estimates of diagnostic accuracy and limit the applicability of the findings. We attempted to address these issues in the quality scoring and sensitivity analyses in KQ 1. We did a systematic check of the included studies to consistently identify the presence of these biases, and include this information in table 10 in the column labeled "criteria for verification by RHC."

Commentator & Affiliation	Section	Comment	Response
Public Reviewer - ACCP 6	Results	Only one prospective study supported the use of biomarker testing with NT pro-BNP combined with echocardiography as possibly helpful in ruling out pulmonary hypertension in symptomatic patients. I believe the reason that there is only one prospective study, is that clinical experience has already proven that this strategy is not a helpful in the clinical arena. BNP levels alone have not been shown to be an accurate surrogate marker for disease severity even in CHF. More importantly, is the echo Doppler cannot be used as a screening tool in asymptomatic or minimally symptomatic individuals. The use of Doppler requires a well-defined tricuspid regurgitation jet which will occur only in patients with elevated pulmonary artery pressure and right ventricular dilation, (i.e. patients with relatively advanced disease). Asymptomatic and minimally symptomatic patients often do not have significant tricuspid regurgitation. They also found that the presence of a pericardial effusion was associated with increased mortality, but this observation has been made previously, and primarily refers to patients with scleroderma.	We have amended the discussion of echocardiographic testing, particularly about screening asymptomatic persons at high risk, to describe this issue. In the studies we evaluated, we attempted to estimate the proportion of patients who failed to have a measureable TR jet; however, this statistic was often unreported. Also, many studies used the presence of a measureable TR jet as an inclusion criterion and failed to detail the number of patients excluded for this reason.
Public Reviewer - ACCP 6	Results	The assessment of drug efficacy presents another constellation of challenges. Virtually all of the randomized clinical trials conducted for this disease were very short duration, 3 to 4 months, and utilized 6 min. walk as the primary endpoint. Because of the lax regulatory requirements for drug approval in PH, most did not include hemodynamic outcome data, or data on mortality. Head-to-head trials have not been conducted, and thus there is little possibility of being able to make judgments on relative efficacy. Thus it is no surprise that their study of comparisons between agents were inconclusive. Combination studies have not been conducted in a structured way since it would require the cooperation and support from two or more pharaceutical companies.	We agree with the reviewer's assessment.
Public Reviewer - ACCP 8	Results	KQ1: The approach has been to look at studies which have assess echo parameters or biomarker or both to identify patients with PAH. It appears that all the studies which included patients with suspected PAH based on abnormal echo, symptoms or those who "screened high risk population" were evaluated together. The question asked was if one of those parameters would identify patients with PAH. The data from very varying disease states Scleroderma, IPAH, sickle cell, portopulmonary are assessed together and I am not sure that is the right approach. I also wondered if it were possible to look at composite of echo findings to predict PAH	We attempted to call attention to differences by diagnostic group and sensitivity analysis by prevalence as a surrogate for the degree of verification bias. Although we had hoped to find data on the predictive and diagnostic value of a composite echocardiographic finding, the studies we identified did not analyze data in such a way as to assess composites (combinations of findings).

Commentator & Affiliation	Section	Comment	Response
Public Reviewer - ACCP 8	Results	KQ1: The other pertinent question would in which high risk patient is screening effective (i.e. preclinical prevalence, sensitivity, cost)	We did not consider cost or cost-effectiveness as an outcome; this was beyond the scope of the project as developed with the technical expert panel.
Public Reviewer - ACCP 8	Results	KQ2: Non invasive assessments echo parameters and or biomarkers were looked at to assess intermediate and long term outcomes – lung transplant/death. And severity of disease using PVR, 6 min. walk, functional status as comparators	Thank you for your comment.
Public Reviewer - ACCP 8	Results	KQ2: Heterogeneous disease states are assessed together. This may have altered the conclusions derived regarding the biomarkers	We added text related to the propensity of most studies to include patients with multiple different etiologies for Category I PAH as one group which complicates interpretation of the findings.
Public Reviewer - ACCP 8	Results	KQ2: Pericardial effusion is strong predictor of mortality – all pericardial effusions?	Studies reporting "presence of pericardial effusion" were included, so the comparison is HR for pericardial effusion versus no pericardial effusion. We did not evaluate results for degrees of pericardial effusion.
Public Reviewer - ACCP 8	Results	KQ2: I am unable to understand the Tables 17, 18	We have modified the table titles for clarity, as follows: Studies reporting changes in mean values over time (Table 18) Studies reporting changes in median levels over time (Table 19)
Public Reviewer - ACCP 8	Results	KQ3: The question asked is the comparative effectiveness of mono vs combinations therapy.	Thank you for your comment.
Public Reviewer - ACCP 8	Results	KQ3: Unclear why the original epoprostenol trial was left out	The 1995 cutoff was chosen based on the FDA approval of Flolan in 1996 as the first of the newer vasodilator treatment for PAH. Unfortunately, this was not early enough to capture the first trial by Rubin et al., 1990. As a result of reviewer's comments, we have revised the searches for KQ 3 to include 1990–1994 and screened the resulting citations. Consistent with the reviewer's suggestions, the only additional study that meets our inclusion criteria is Rubin. We amended the protocol to reflect this change. The study by Rich et al. was ineligible for inclusion because it did not include a comparator group. The study by Barst et al. was included as extension study of the RCT by Rubin et al., 1990.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer - ACCP 8	Results	KQ3: When looking at the drug effect all studies in varied population are included (i.e. for bosentan we have the BREATH-1, the PREATH 5, ASSEST)	Thank you for your comment.
Public Reviewer - ACCP 8	Results	KQ3: Prostacyclins are clubbed together as a class while they are different agents and there are difference in the inhaled and the parenteral – for efficacy/SE etc	We have made the following revisions to the report: (1) Prostanoids administered by an inhaled route are identified in each forest plot, and (2) we report a separate summary measure associated with inhaled prostanoids for each outcome for which both inhaled and noninhaled routes were studied.
Public Reviewer - ACCP 9	Results	p. 70: "…indicating that changes in BNP were not associated with…" – "changes in" is not clear. Changes over time? Changes from a baseline value? Changes from a normal range (i.e., an abnormal value)?	At the beginning of the section on responsiveness of biomarkers and echocardiography we consider change from baseline over two or more different time points. In the previous sections on prognosis and predictive value of biomarkers, we consider cross-sectional relationships between biomarkers and outcome; we have changed the final report to use the term "differences in [biomarker or echocardiographic parameter]" to refer to these relationships, and reserved the term "change in [biomarker or echocardiographic parameter]" to refer to change from baseline in studying responsiveness.
Public Reviewer - ACCP 9	Results	p. 70: Where is the reader told how "pericardial effusion" is defined? Most prior studies have ranked the effusion, and so one is left to wonder if a "trivial" or "trace" pericardial effusion has the same prognostic implication as a "large" effusion. My apologies if I have missed itm, but it would seem from what is presented here one might (erroneously, I suspect) conclude that trace/trivial effusions are just as important as large ones.	Studies reporting "presence of pericardial effusion" were included, so the comparison is HR for pericardial effusion versus no pericardial effusion. We did not evaluate results for degrees of pericardial effusion.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer - ACCP 9	Results	Table 17 (p74): We are not told the designs of the studies included. They differ quite a bit. Is that meant to be reflected in the summary quality grade provided? As noted above, we are not told what led to downgrading in study quality grade.	We agree that the individual study designs are not neatly described. We struggled with such a description for inclusion in a table but ended up using multiple design features to assess study quality, including direction of inquiry (prospective versus retrospective), basic study architecture (cohort, case-control, nested case-control, convenience sample, case series, etc.). In many cases, key information necessary to characterize the design was not described; in others the assembly of the patients was for another purpose (e.g., RCT of treatment). As described in the Methods, we did consider these issues in the quality rating. As we describe in Methods, individual study quality ratings were evaluated based on features described in Table 3 and grouped in categories good, fair, or poor. Specific ratings for each study are provided in Appendix D along with features that limit the applicability of the findings. (Note that applicability concerns were not given a single rating as study quality was. We made an editorial decision not to present the myriad data elements that went into the quality grade due to space constraints. Quality ratings were reached as the consensus of two (and sometimes three) reviewers of the article; disagreements were resolved by consensus.
Public Reviewer - ACCP 9	Results	Table 17-19 (and associated text): The outcomes reported are not terribly useful. What were the outcomes of the patients? Did they have an improvement in walk distance? In functional class? If you have not clearly defined a valid "surrogate" for an important patient outcome, knowing that a marker changes or does not, is by itself not terribly useful.	We agree with the reviewer; while most studies likely included changes in 6MWD, functional class as well, none specifically correlated these outcomes with changes in biomarker or echocardiography, making the data less useful.
Public Reviewer - ACCP 9	Results	p. 86 Third bullet: "…associated with prostanoids…" – all prostanoids, or just inhaled? This is clinically an important question.	We added a separate analysis of adverse events specific to inhaled prostanoids in the Results section and revised the Key Points regarding adverse events to specify inhaled prostanoids.
Public Reviewer - ACCP 9	Results	p. 86 – no comment regarding comparisons among classes of drugs, or individual agents within a class?	We added the following Key Point: We did not identify any eligible studies that evaluated the comparative effectiveness of calcium channel blockers on intermediate-term and long-term patient outcomes, or that randomized treatment- naïve patients to monotherapy versus combination therapy, or that directly compared two drug classes.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer - ACCP 9	Results	p.96, table 23: Why haven't observational studies of calcium channel inhibitors been included? (e.g.,L Sitbon et al 2005 PMID15939821). Why were certain observational studies of other agents excluded (e.g., McLaughlin et al PMID16919005, Sitbon et al. PMID 16055621)? The list of excluded studies in the appendix does not really make this clear.	In the TEP call of 12/8/11, discussions about calcium channel blockers included the following points: (1) Although sometimes used for PAH, CCBs are not FDA approved for the treatment of PAH and are only effective in a small subset of patients with specific criteria (vasodilator response). (2) CCBs do not appear to have been studied or used in a manner that lends itself to comparative effectiveness analysis and so may not be appropriate for this review. The suggestion was made to discuss CCBs as an acceptable treatment option in the background section. While we retained CCBs in our searches and analysis plan, we did not find any eligible studies. As stated in the report: "We did not identify any eligible studies published after 1995 that evaluated the safety or efficacy of calcium channel blockers on intermediate-term or long-term patient outcomes." The specific observational studies mentioned by the reviewer were considered but not analyzed: - Sitbon, 2005 PMID: 15939821—was described in background section and cited in the report. - Sitbon 2005 PMID: 16055621. Was excluded and is listed in Appendix E. - McLaughlin 2006 PMID: 16919005. Was excluded and is listed in Appendix E.

Commentator & Affiliation	Section	Comment	Response
			(Continued from previous table cell) These last two studies had peculiar features in assembling comparison groups that resulted in exclusion. Sitbon cobbled together the bosentan group from treatment groups of two placebo-control trials that are included in our analysis of RCTs. McLaughlin similarly assembled disparate groups of patients from RCTs, open-label extensions of RCTs and retrospective registry data. Note that for clarity, we reorganized the list of excluded studies in Appendix E in alphabetical order by first author, accompanied by the reason for exclusion.
Public Reviewer - ACCP 9	Results	p.98: "…we would infer efficacy of tadalfil from controlled trials of tadalafil both with and without background therapy." The meaning here is not clear, but worrisome if I have interpreted correctly (I hope not). It is also not clear where such data have been incorporated. It seems the authors are saying that they included in the meta-analysis of a drug's efficacy data from combination studies comparing the addition of that drug to some "background" that included another PAH drug. (i.e. lumping de novo tratment data with add-on treatment data for a given drug). That assumes (dangerously) that if each of the PAH-specific drugs being discussed have efficacy, that their efficacies are purely additive and independent. If, for example, drug A added to placebo results in an improvement of X, and when added to drug B (instead of placebo) results in an improvement of 0.5X, would the authors combine these data to conclude that the best estimate of drug A's overall affect is 0.75X? It would be naïve and foolish to assum that a patient already treated with one agent (even of a different class) would necessarily have the same response to another class as if s/he had not already been treated for the disease in some manner. I am sure I must be misunderstanding the authors' intent.	We appreciate the reviewer highlighting this important issue. We have revised the text to clarify that our approach assumes independent and additive effects of the experimental drug relative to any or all of the other background therapies received by patients enrolled in the trial (including, but not limited to, other PAH-specific drugs, supplemental oxygen, vasodilators, etc.). In effect, all of the RCTs included in this report evaluated the efficacy associated with the addition of a single drug to a combination therapy regimen. For the purpose of this report, however, we use the term "combination therapy" to refer to studies that included a prostanoid, a PDE inhibitor, or an endothelin antagonist as part of the background therapy. We explicitly mention the assumption of additive effects that this analysis requires, and describe some data from PHIRST study that would tend to contradict this assumption.
Public Reviewer - ACCP 9	Results	p.101 – top: No data for inhaled treprostinil?	We have revised the forest plots to indicate that the TRIUMPH-1 study by McLaughlin et al. (2010) used aerosolized treprostinil.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer - ACCP 9	Results	p.101 – bottom: It seems inappropriate to combine these data to assess mortality in assessing combination vs monotherapy given the very clear heterogeneity among the studies (different drugs used!). This reader is left wondering by what criteria the "little study heterogeneity" was established (p.12). Same concern on p 103 regarding combining the studies of different agents in an assessment of combination vs monotherapy on hospitalization (p 103).	We agree that the differences between drugs severely limit the usefulness of this analysis; we believe we provided appropriate caveats to its interpretation. "Little heterogeneity" refers to the degree of between-study heterogeneity in the results as measured by the I-squared statistic. Note that while older methods for testing for homogeneity (chi-square) are limited in the ability to detect heterogeneity when the number of studies is small as it is in this case. We have removed the phrase "little study heterogeneity" in the final report and tried to clearly distinguish between statistical assessments of heterogeneity of results versus clinical heterogeneity between trials.
Public Reviewer - ACCP 9	Results	P 101 – you should indicate whether you are able to compare the results of endothelin receptor antagonists and phosphodiesterase inhibitors on 6 MWD (likely no, but important to say so).	The revised report includes a sufficient number of studies to generate a summary estimate for change in 6 MWD for each of the three drug classes.
Public Reviewer - ACCP 9	Results	This section on 6MWD is one of many examples of very mechanical reporting that fails to synthesize the information in a way that is useful for clinicians.	We have revised the forest plots and text to indicate more clearly the findings from each individual study and the meta-analysis results for each drug class for each outcome, including 6 MWD.
Public Reviewer - ACCP 9	Results	Figure 26: PHIRST 1 has been listed as the background therapy as "none" – not true. Some patients were on bosentan!	We appreciate this comment. We have revised the text to clarify that 53% of patients in the PHIRST study were on bosentan, and we summarize the findings from the published subgroup analyses. For the purpose of classifying studies as monotherapy versus combination therapy, however, we believe that 53% is not sufficiently high to classify the PHIRST study as primary a combination therapy study.
Public Reviewer - ACCP 9	Results	P106: last paragraph: "a significant improvement in cardiac index of -1.00…" A reduction in CI is not an improvement. I am thus left confused by figure 31. Channick et al 2001 (Study 351) showed an improvement of 0.5 (or 1.0 "corrected for placebo). You have either misinterpreted the data or mis-transcribed it.	The Channick, 2001, study abstract states "The cardiac index was 1.0 L min(-1) m(-2) (95% CI 0.6-1.4, p<0.0001) greater in patients given bosentan than in those given placebo" The draft report reversed the data for this study. We have corrected this error in the revised report.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer - ACCP 9	Results	Adverse events: You state you were unable to combine the data for each of the AEs (headache, peripheral edema, etc) when assessing combination versus monotherapy because of high heterogeneity. I do not argue the point. But, you use pretty much the same studies (and combine their data) when assessing combination versus monotherapy effects on 6 min walk distance. The differing rationales are not clear.	6MWD is a clearly defined continuous outcome, whereas adverse events are discrete and relatively rare events. We believe that there are sufficient data to estimate mean differences in 6MWD for the combination versus monotherapy comparison. In the case of adverse events, the two or three studies that reported discrete adverse events did not define or collect them consistently across studies. Therefore, although the outcomes were drawn from the same studies, we felt meta-analysis of adverse events would not be appropriate.
Public Reviewer - ACCP 9	Results	P112 and p113: there needs to be consistency in the use of terms such as “demonstrated that…” versus “…suggests that…” One needs to be exceedingly cautious in language that makes stronger inferences if not based on higher quality data.	We revised the report by using the term “demonstrated” only when there was sufficient strength of evidence to support a treatment effect.
Public Reviewer - ACCP 9	Results	Table 24: no mention of tadalafil?? Was this not assessed? If not, the SR/MA is already out of date. For treprostinil, (in this table and in general) is seems clinically inappropriate to combine inhaled, subcutaneous and intravenous unless there are good data to indicate one should do so. Assuming they are equally efficacious is dangerous, and requires proof.	We did not include tadalafil in an SOE table for mortality in the draft report because we did not, at that time, include comparisons in the SOE table for which there was only a single eligible study. We have revised our approach, and have included each of the individual drugs in the SOE table for mortality, irrespective of the number of studies. We have revised the text and tables to clarify that only subcutaneous and IV treprostinil were studied.
Public Reviewer - ACCP 9	Results	Table 22: see remark above regarding defining/characterizing pericardial effusions.	Studies reporting “presence of pericardial effusion” were included, so the comparison is HR for pericardial effusion versus no pericardial effusion. We did not evaluate results for degrees of pericardial effusion.
Peer Reviewer 1	Discussion	The authors correctly state that there are significant research gaps in terms of managing and treating PAH.	Thank you for your comment.
Peer Reviewer 1	Discussion	The authors might also mention that comparative effectiveness of serial evaluation of individuals with PAH has not been significantly studied. Whether echocardiography alone without routine cardiac catheterization and laboratory tests can be used to follow individuals with PAH has not been studied.	Our literature search update yielded a study of serial echocardiography for screening in systemic sclerosis. We did not find data on serial echocardiography for monitoring patients diagnosed with PAH.
Peer Reviewer 1	Discussion	Additional echocardiographic measurements such as end diastolic pulmonary regurgitation gradients, mean tricuspid regurgitation gradient, and Doppler tissue imaging of the tricuspid annulus require more evaluation and comparison among each other among individuals with PAH.	We have added a point to the Research Gaps for KQ 1, as follows: Consider further studies of additional promising measures such as end diastolic pulmonary regurgitation gradient, mean tricuspid regurgitation gradient, and Doppler tissue imaging of the tricuspid annulus.
Peer Reviewer 3	Discussion	The implications of the results are described in the discussion.	Thank you for your comment.

Commentator & Affiliation	Section	Comment	Response
TEP Member 1	Discussion	Yes. Of course the major limitation is that, despite a great deal of work in this field over the past 15 years, PAH remains a relatively uncommon condition and the strength of recommendations reflects the limited long-term data and the small population studied. Since these are unlikely to change, any methodological ideas to enhance the possible conclusions of future studies would be welcome.	We have included a list of suggestions for future research needs.
TEP Member 2	Discussion	I think this is where the main limitations are. Obviously, there were questions that were not answered here. There were attempts to glean data, but the clinical relevance is questionable. There are some misinterpretations or inaccuracies. The second key point in KQ1 indicates that echo shows good accuracy with RHC for PASP. But, it specifies PAH which is probably not accurate. Just because the estimated PASP on echo correlates with the measured PASP on cath, it doesn't mean the patient has PAH. There are many patients who are worked up for PH based on an elevated PASP on echo who end up having left heart or lung disease. Based on the key points for KQ1, really very little is added to the current knowledge base. Similarly, in the key points for KQ2, correlations are made, eg, PASP on echo and RHC, but it is not translated to anything that is clinically meaningful. Does that really change management. In fact, most of what is discussed in KQ2 does not add at all to clinical practice. The research gaps could use some refining. In fact, many of the studies they are suggesting would be a low level of evidence based on trial design. For example, post hoc sub group analysis of treatment efficacy by WHO FC or etiology would be hypothesis generating, not definitive at all. Others are simply not feasible, such as large or long term head to head trials.	We agree that the correlation between echo sPAP and RHC sPAP is not sufficient for diagnosis of PAH, but it has important implications for the use of echocardiography in screening and case finding. We have mentioned the limitations in assessing left heart filling pressures by echocardiography. We agree that the correlations and hazard ratios and other associations described for KQ 2 have little clinical utility, failing to rise to the level of therapeutic efficacy; however, we hope that noting the shortcomings in the evidence will lead to future studies to address this lack.
TEP Member 3	Discussion	The literature search for the evaluation of different biomarkers had no possibility of being helpful since the only biomarker that has been systematically evaluated has been BNP. None of the RCTs, which were industry sponsored, included any substudies to allow for these types of assessments. Clinical practice has already shown that BNP has limited utility in the diagnosis or treatment of these patients. Their findings support this.	Thank you for your comment.

Commentator & Affiliation	Section	Comment	Response
TEP Member 3	Discussion	With respect to comparing monotherapy with combination therapy, the limited clinical trials including combination therapy, and the selection of the 6 min. walk as the primary outcome variable precludes any meaningful conclusion. Also, for the reasons stated above, it was predictable that the evidence would be inconclusive regarding a reduction in mortality with monotherapy or combination therapy.	Thank you for your comment.
TEP Member 3	Discussion	The biggest flaw with this project has is that it has considered patients with PAH as having one disease. Nothing could be further from the truth. Whilte patients with idiopathic pulmonary hypertension have an insidious onset of symptoms, and a subset of these patients may have a genetic mutation placing them at high risk, they really represent a great minority of the group. Patients with connective tissue diseases, primarily scleroderma, represent the largest subset of patients, and all of these patients tend to be symptomatic early in their disease related to varying degrees of lung involvement. Since the initial symptom of PH is dyspnea, which is the initial symptom of pulmonary fibrosis, there is no way to know if a scleroderma patient has one or the other, or both. Clinical studies in scleroderma patients have suggested that a reduction in diffusing capacity from pulmonary function testing tends to be a useful biomarker to identify those patients most likely to develop pulmonary hypertension. However pulmonary hypertension associated with scleroderma is very different than idiopathic pulmonary pretension in its prevalence, onset of symptoms, and respons to therapy, and survival. The same can be said for patients with congenital heart disease and pulmonary hypertension. It is illogical to look for a biomarker to detect early onset of pulmonary pretension in patients with congenital heart disease since they will be symptomatic from birth. Additionally their survival is dramatically better than patients with connective tissue disease or idiopathic PH.	Most studies included patients with multiple etiologies of Category I PAH as one group making it more difficult to draw conclusions from the data. We added language in the recommendations for future research highlighting the need for studies which separate out the different etiologies of Category I PAH, although this will likely be difficult given the low prevalence of the disease in general, making it difficult to garner an adequate number of patients for many of the subgroups.
TEP Member 3	Discussion	The conclusion in the report that the findings do not support any recommendations for replacing existing measurement tools to assess disease severity prognosis or respons to therapy was anticipated by this reviewer from the outset.	Thank you for your comment.
TEP Member 4	Discussion	The implications ar clearly stated with limitations identified.	Thank you for your comment.
TEP Member 4	Discussion	As for the statement, “For KQ 2, our findings support using echocardiography or biomarkers in place of existing measurement tools to assess disease severity, prognosis, or response to therapy.” I will defer to other Reviewers’ comments.	Thank you for your comment.

Commentator & Affiliation	Section	Comment	Response
TEP Member 4	Discussion	The future research section is clear and marks important gaps to address. The gaps identified are consistent with gaps identified I other settings, such as NHLBI workshops, working groups, etc., particularly in the need to perform different types of interventional stuides in the future.	Thank you for your comment.
TEP Member 4	Discussion	The research section presents the gaps in a way that should be easily translated in the future.	Thank you for your comment.
TEP Member 5	Discussion	Reasonable.	Thank you for your comment.
TEP Member 6	Discussion	There must be more attention on the effect of the referred population. Results from many studies are skewed since they come from PH centers and include strict inclusion criteria. Diagnostic and prognostic performance will change markedly by broadening to the general population. PASP by echo is meaningful only when RV function and/or stroke volume is considered. The FP rate will be high if one does not consider that PASP can increase to >40 mm Hg in normal individuals during high output states. On the other side of the spectrum, low-output low pressure PH is an entity that, according to your analysis, may need addressed.	We discuss skewing of the patient population in our Applicability section, and we have also included language in our Background section to acknowledge the point about stroke volume vis a vis threshold for sPAP.
TEP Member 7	Discussion	The implications are clearly stated and the limitations are adequately reviewed. However, much of the discussion and conclusions are lost in the length of the overall text and would be better discussed earlier in the report, including the summary section ES-1 thru ES-28. Page 119 KQ 1: Screening for PAH I agree with the first part of the statement “In low prevalence populations (<10%), negative predictive value of a normal sPAP is high, suggesting that echocardiography with a low threshold may be an appropriate test in asymptomatic high-risk populations” However, I think that there is insufficient evidence that a normal sPAP on echo can exclude PAH “in patients with symptoms suggesting PAH” and that this above statement seems to contradict the paragraph that follows it in which it is well stated that “echocardiographic estimates of sPAP often over- or under-estimate pulmonary artery pressure enough to result in misclassification according to PAH diagnostic threshold… echocardiography cannot be relied upon to exclude pulmonary hypertension if pretest probability is high.”	We agree that the report is lengthy. We are working within an organizational structure common to Comparative Effectiveness Review reports, including lengths limits on the Executive Summary. We do not agree that much of the main report’s discussion would fit into the Executive Summary. We have expanded the discussion of misclassification data, adding data from 3 new studies (Table 12, page 54). Interestingly, the additional data indicated somewhat lower standard deviations than the REVEAL study, suggesting that the misclassification problem might be slightly smaller than REVEAL suggests. We also followed the comment about using a normal sPAP on echo to exclude PAH in symptomatic patients with several important caveats.
TEP Member 8	Discussion	Excellent section	Thank you for your comment.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer - ACCP 3	Discussion	*"Currently, right heart catheterization (RHC) is the gold standard for diagnosing and monitoring progression of PAH"* I agree with the "gold standard for diagnosing", but I am not sure whether I consider RHC the "gold standard" for "monitoring progression" of PAH. It can certainly contribute, and I suppose many would consider it the "gold standard", but my current gold standard is multimodal, including clinical assessment by history and physical examination, 6 minute walk, echo, biomarkers e.g. BNP.	We chose to focus on echocardiography and biomarkers because there is particular uncertainty about the extent to which these could substitute for the reference standard of RHC in the monitoring process.
Public Reviewer - ACCP 6	Discussion	The literature search for the evaluation of different biomarkers had no possibility of being helpful since the only biomarker that has been systematically evaluated has been BNP. None of the RCTs, which were industry sponsored, included any substudies to allow for these types of assessments. Clinical practice has already shown that BNP has limited utility in the diagnosis or treatment of these patients. Their findings support this.	Thank you for your comment.
Public Reviewer - ACCP 6	Discussion	With respect to comparing monotherapy with combination therapy, the limited clinical trials including combination therapy, and the selection of the 6 min. walk as the primary outcome variable precludes any meaningful conclusion. Also, for the reasons stated above, it was predictable that the evidence would be inconclusive regarding a reduction in mortality with monotherapy or combination therapy.	Thank you for your comment.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer - ACCP 6	Discussion	The biggest flaw with this project has is that it has considered patients with PAH as having one disease. Nothing could be further from the truth. Whilte patients with idiopathic pulmonary hypertension have an insidious onset of symptoms, and a subset of these patients may have a genetic mutation placing them at high risk, they really represent a great minority of the group. Patients with connective tissue diseases, primarily scleroderma, represent the largest subset of patients, and all of these patients tend to be symptomatic early in their disease related to varying degrees of lung involvement. Since the initial symptom of PH is dyspnea, which is the initial symptom of pulmonary fibrosis, there is no way to know if a scleroderma patient has one or the other, or both. Clinical studies in scleroderma patients have suggested that a reduction in diffusing capacity from pulmonary function testing tends to be a useful biomarker to identify those patients most likely to develop pulmonary hypertension. However pulmonary hypertension associated with scleroderma is very different than idiopathic pulmonary pretension in its prevalence, onset of symptoms, and respons to therapy, and survival. The same can be said for patients with congenital heart disease and pulmonary hypertension. It is illogical to look for a biomarker to detect early onset of pulmonary pretension in patients with congenital heart disease since they will be symptomatic from birth. Additionally their survival is dramatically better than patients with connective tissue disease or idiopathic PH.	Most studies included patients with multiple etiologies of Category I PAH as one group making it more difficult to draw conclusions from the data. We added language in the recommendations for future research highlighting the need for studies which separate out the different etiologies of Category I PAH, although this will likely be difficult given the low prevalence of the disease in general, making it difficult to garner an adequate number of patients for many of the subgroups.
Public Reviewer - ACCP 6	Discussion	The conclusion in the report that the findings do not support any recommendations for replacing existing measurement tools to assess disease severity prognosis or respons to therapy was anticipated by this reviewer from the outset.	Thank you for your comment.
Public Reviewer - ACCP 9	Discussion	P123: last paragraph: "…more broadly, this is…alone)." Your meaning is unclear. If anything, one might have though the opposite, that more generally this is bosentan/sildenafil plus prostenoid vs bosentan/sildenafil alone.	We have clarified the sentence by removing the parenthetical statement.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer - ACCP 9	Discussion	Applicability: "may not be applicable to screening of asymptomatic patients." My apologies, but there is no other kind of screening, by definition. This requires attention throughout to avoid embarrassment (e.g., in your conclusion, "when screening symptomatic patients…")	The type of screening we are interested in is sometimes called case finding. Aimed at populations that are at high-risk, one would still expect them to be free from signs or symptoms of the target condition. Many studies of high-risk populations did not exclude patients with signs or symptoms, lumped together asymptomatic with symptomatic individuals or tested populations whose symptoms were attributed to a disease other than the target condition. We believe that although these studies do not conform to a strict definition of screening, they nevertheless provide information that is applicable to this situation.
Public Reviewer - ACCP 9	Discussion	P 128 bottom: are you referring of the ESC/ERC guidelines through this discussion ("This guideline recommends…")?	Yes, the European Society of Cardiology (ESC)/European Respiratory Society (ERS).
Public Reviewer - ACCP 9	Discussion	P129: on what basis have you concluded that your data are generally consistent with monotherapy first, and combination later? I'm not saying I disagree, but I'd like to know how the data here have shown that.	We have clarified the discussion to state that the findings from our meta-analyses of the few studies that compared combination therapy with monotherapy suggest, but do not prove, that combination therapy confers more benefit than does monotherapy in the treatment of PAH. We now describe the principal caveat to these data; namely that they come from addition of a second drug to existing treatment compared to continued existing treatment rather than a head to head comparison of combination versus monotherapy in treatment-naïve patients.
Public Reviewer - ACCP 9	Discussion	P129: "… we anticipated better quality data…" Unfortunately, most people in this field did not anticipate better quality data.	Thank you for your comment.
Public Reviewer - ACCP 9	Discussion	P130: "Also, there are no clear data on benefits of early treatment…" – although only one study, it is incorrect to say there are no clear data (Galie 2008, PMID18572079)	This section is discussing screening. Asymptomatic patients diagnosed through echocardiographic screening would be at an even earlier point in disease than the WHO FC II studied by Galie et al., 2008.
Public Reviewer - ACCP 9	Discussion	P132: "data on minimum clinically important change in 6MWD. See Gabler et al (PMID 22696079) – published since you did your review.	Thank you for bringing this study to our attention. In the revised report, we put the 6MWD findings (improvement of 24 meters for combination versus monotherapy) in context by citing and discussing the recently published study by Mathai et al. that estimates the MID for the 6MWD among patients with PAH to be approximately 33 meters.
TEP Member 5	Appendixes	Fine…	Thank you for your comment.
TEP Member 6	Appendixes	No comments.	Thank you for your comment.
TEP Member 6	Abbreviations	Pretty heavy – it would be very helpful to have a key of abbreviations used.	An Abbreviations table is provided at the end of the report.
TEP Member 5	Tables	Fine except shading might require a 508 compliance notation.	Shading question to be addressed by AHRQ.

Commentator & Affiliation	Section	Comment	Response
TEP Member 6	Tables	Very labor intensive but very effective	Thank you for your comment.
Public Reviewer - ACCP 1	Tables	More specifically, regarding Evidence Table 24, it is nice that they brought together the studies in a systematic manner, but the meta-analysis here is not appropriate. The PACES study is the most methodologically rigorous, and the plot should suggest that.	Weighting in the forest plot is by inverse variance weighting (essentially proportional to sample size). The PACES study (Simonneau et al., 2008) had a proportionally large effect on the summary estimate (with the summary estimate being very similar to that reported by Simonneau et al., 2002) Nearly all of the RCTs included in the KQ 3 analyses were rated as good quality, and none was rated as poor. We did not, therefore, conduct sensitivity analyses that excluded the few fair-quality studies.
TEP Member 5	Figures	Fine although the PRISMA might be expanded so studies used quantitatively versus qualitatively might be broken out.	Although the Literature flow diagram (Figure 3) does not detail the specific analyses that included articles are for, this information is described in the Results section.
TEP Member 6	Figures	Also very effective and clear.	Thank you for your comment.
Public Reviewer - ACCP 1	Figures	Regarding Iloprost, the studies consistent with each other to the point where we can compare numbers in a plot. There are huge confidence intervals that make everything look insignificant, and it does not make sense clinically to put these studies together because it is not usable.	For each forest plot, we included all of the eligible studies that reported a given outcome, irrespective of the width of the confidence intervals of any given study.
Public Reviewer - ACCP 1	Figures	Regarding Bosentan (Figure 20), the plot includes the Barst study, which has a totally different patient population than the others. In fact, all three studies included were different. The Channick study should be combined with the Rubin study in the plot, but not the Barst study. These study inclusion issues may be why the summary statistics do not look very good.	We appreciate the reviewer's comment and concern. We discuss in the Limitations section the limitations introduced by study heterogeneity, including heterogeneity in patient populations.
Public Reviewer - ACCP 1	Figures	Regarding Figure 23, it makes sense to combine, but the dosing needs to be looked at further, as they are different across studies. The doses of 5mg should be combined. Also, the studies are geographically different. At the very least, combine the active doses. Also, the report concluded that the Channick study in Lancet was an improvement, which may be a wrong conclusion.	The estimates in Figure 23 combine the active doses within each study. We added an explanation about how multiple doses were analyzed. We have revised the Table 23 to clarify that ARIES 1 and ARIES 2 were conducted in different geographical regions. Multidose studies included only a single control group; we therefore did not report each dose individually in the forest plots to avoid double counting patients. We did not combine single doses (e.g., 5 mg) because different multidose studies had different control and comparison groups. The Draft report erroneously reported that in Channick et al., 2001, the between-group difference in cardiac index as -1.0 (representing an improvement) instead of 1.0 (representing worsening). We have corrected that error.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 1	General: Quality of the report	Good	Thank you for your comment.
Peer Reviewer 2	General: Quality of the report	Good	Thank you for your comment.
Peer Reviewer 3	General: Quality of the report	Superior	Thank you for your comment.
TEP Member 1	General: Quality of the report	Superior	Thank you for your comment.
TEP Member 2	General: Quality of the report	Good	Thank you for your comment.
TEP Member 3	General: Quality of the report	Fair	Thank you for your comment.
TEP Member 4	General: Quality of the report	Good	Thank you for your comment.
TEP Member 7	General: Quality of the report	Superior	Thank you for your comment.
TEP Member 8	General: Quality of the report	Superior	Thank you for your comment.
Peer Reviewer 1	General: Clarity/ usability	I believe the results are clearly presented and that the conclusions highlight a need for future research.	Thank you for your comment.
Peer Reviewer 2	General: Clarity/ usability	See comments above, The accuracy and presentation can be improved substantially. The document is useful to health care providers and researchers. It will help with some clinical decisions, but not with many others due to insufficient data. It also points out many knowledge gaps, although the recommended avenues for future research do not consider the practical limitations faced by the field.	Thank you for your comment.
Peer Reviewer 3	General: Clarity/ usability	The report is well-structured and organized. The document will be a useful guide for clinicians, researchers, and policy makers.	Thank you for your comment.
TEP Member 1	General: Clarity/ usability	Yes, although I do have a concern that the weakness of conclusions regarding screening high-risk populations and outcomes of therapy may be used by third-party payors to dictate the least expensive approaches, even if the clinical recommendation suggests, albeit to strongly, otherwise.	While payment policies are an important concern, we believe that the availability of complete and unbiased data is a good foundation for discussion and decisionmaking. We comment in the Discussion section on Implications for Clinical and Policy Decisionmaking: "The lack of direct comparisons between assessment strategies, and the lack of measures of clinical outcomes associated with screening diagnostic or prognostic testing would not seem to support more directive recommendations regarding testing modalities."
TEP Member 2	General: Clarity/ usability	Yes, as best as can be expected with such a data dense report.	Thank you for your comment.

Commentator & Affiliation	Section	Comment	Response
TEP Member 3	General: Clarity/ usability	This report follows the characteristics of other AHRQ reviews. It is very well organized, easy to read and will structured. It is thorough in its approach and provides a comprehensive overview addressing the questions posed.	Thank you for your comment.
TEP Member 4	General: Clarity/ usability	The report is well-organized and main points are clearly presented.	Thank you for your comment.
TEP Member 4	General: Clarity/ usability	It is unclear whether the conclusions will have an impact on clinical practice decisions, but this is a function of the limitations encountered during the analysis rather than the outcome of the analysis itself.	Thank you for your comment.
TEP Member 7	General: Clarity/ usability	The report is well organized, but the structure is somewhat complicated. This is an exhaustive report that does not lend itself well to easy perusal or quick referencing. The summary statement ES-1 thru ES-28 is helpful, but repeating much of this information in the introduction (pages 1-15) seems unnecessary and confusing to the reader. A better approach may be to keep the summary document and then refer the reader to more detailed review of the results and discussions.	Thank you for your comment.
TEP Member 7	General: Clarity/ usability	The main points are clearly presented. However, it is doubtful that the conclusions can be used to inform policy or practice decisions. For example, I believe the following statement (1st paragraph, page 129) is correct: “Echocardiography-derived sPAP shows promise as a possible surrogate marker for RHC-sPAP, but whether or not this measure alone is adequate to assess disease severity, prognosis, or response to therapy is unclear, and so this evidence is insufficient to support recommendations regarding policy changes in regard to this measurement tool.”	Thank you for your comment.

Commentator & Affiliation	Section	Comment	Response
TEP Member 7	General: Clarity/ usability	Other major limitations to influencing policy/or practice decisions is described on page : “Second, we anticipated better quality data for the questions about screening and diagnosis (KQ 1) and prognosis (KQ 2) than we actually found” For KQ 3, “there is the potential for bias due to analyzing published studies, which are more likely to have positive results.” This later point is particularly important in that the major studies reviewed were designed, sponsored and conducted by industry. Publication of all the data from these studies were industry controlled. The above limitations are significant and greatly alter the impressions put forth by the key messages. Rather than burying these limitations in the conclusion section that occurs after 100 pages of text, they should be brought forward to where the key messages are first presented.	Publication bias is assessed more formally in the final report. Key messages are intended to relate to findings, but not potential limitations.
TEP Member 7	General: Clarity/ usability	The final section on Research Gaps is excellent and should be mentioned earlier in the report. This may in fact be the greatest value of the report. Previous studies have been designed to examine short term efficacy of particular treatments. Numerous investigators continue to study many of the questions raised in this report, but in an uncoordinated manor. The small number of patients with this disease greatly limits the number of studies that can be done. It is extremely helpful to the research community to have a list of established goals that should be addressed in future studies. This list should be made available to academic as well as industry related investigators.	Thank you for your comment.
TEP Member 8	General: Clarity/ usability	Organization excellent	Thank you for your comment.
Peer Reviewer 1	General	The report reviews many articles on the screening, prognosis, and treatment of pulmonary arterial hypertension (PAH). The report is well written and provides good organization of the topic and references.	Thank you for your comment.
Peer Reviewer 1	General	The authors conclude that echocardiography and NTpBNP are accurate in screening for PAH, but that further research is needed regarding the prognosis and treatment of PAH.	Thank you for your comment.
Peer Reviewer 1	General	The authors adequately point out the limitations of the evidence based used in their study and research gaps regarding evaluation and treatment of PAH.	Thank you for your comment.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 1	General	While the authors specifically analyzed studies on PAH, the screening tests described (especially BNP, NTpBNP, and multiple echocardiographic parameters) have been valuable in measuring hemodynamics among patients not specifically chosen for evaluation of PAH. This focus on PAH omits valuable literature on other parameters including mean tricuspid regurgitation gradient, and diastolic regurgitation gradient, and tricuspid Doppler tissue E/e^1 that are useful in evaluating pulmonary artery pressure.	Given the large number of different biomarkers and echocardiographic parameters that have been evaluated, review of each was impractical. We developed an abbreviated list of those that had been studied more often using several techniques including reference to guidance documents (e.g., ASE guidelines), consultation with TEP members, and preliminary literature screening. The specific echo parameters mentioned (mean tricuspid regurgitation gradient, diastolic regurgitation gradient, and tricuspid Doppler tissue E/e^1) are less often studied, and not recommended in ASE guidelines.
Peer Reviewer 2	General	This is a thorough systematic review of the literature related to the 3 key questions posed by the analysts. The questions are important, the analysis seems appropriate and novel conclusions are drawn, especially with regard to screening and management. This document will be useful to the field, especially with regard to pinpointing knowledge gaps. Nonetheless, there are a number of improvements that could be made in presentation and improving the accuracy of some of the conclusions, especially those re adverse side effects as detailed below. Also, there is no mention of slowing the clinical deterioration as an outcome, which is increasingly used in more recent trials. This seems to be an odd omission. Please explain.	We revised the forest plots and associated text for all adverse events, including cough, to generate separate estimates for all prostanoids and inhaled prostanoids. We revised the discussion to include the limitation associated with the limited, unstandardized, and nonsystematic reporting of adverse events. The clinical deterioration outcome is not yet widely used and was not a feasible choice for this retrospective review. We did consider individual outcomes that would indicate clinical deterioration including hospitalization (analyzed), need for transplantation (not enough data to be analyzed), mortality (analyzed), 6MWD (analyzed), and others.
Peer Reviewer 3	General	This is a useful and clinically meaningful report. The key questions were explicitly stated.	Thank you for your comment.
TEP Member 1	General	This is a comprehensive and balanced evidence-based analysis of the state of early diagnosis and management of PAH which covers these issues completely and provides recommendations for further study. While many of the conclusions and recommendations are not strong, they reflect the state of knowledge of this rare disease. The target population would be practitioners wishing guidance for management, third-party payers, and government.	Thank you for your comment.
TEP Member 2	General	As with any report such as this, it is mostly a “data dump.” There are several areas in which the clinical relevance is questionable. In some cases, the key questions are impracticle.	Thank you for your comment.

Commentator & Affiliation	Section	Comment	Response
TEP Member 3	General	This project had as one of its initial goals an evaluation of the validity reliability and feasibility of echocardiography and biomarker testing for the screening diagnosis and management of PAH. Secondly they sought to determine whether the use of echocardiography or biomarkers would affect clinical decision-making and clinical outcomes. Finally they sought to determine which medications are effective for treating PAH and whether combination therapy is more effective than monotherapy. While AHRQ methodology has a proven track record for conducting exhaustive and accurate assessments of published data referable to diseases, I fear that this project was futile from the outset. It is unclear to me why the goals of the project were chosen, but in my opinion had they consulted the leading experts in the field this project would not have been done. The single largest flaw is the assumption that all of the different forms of pulmonary hypertension listed under the category of PAH are similar. The reality is they are very different. Secondly, the pulmonary hypertension community and regulatory authorities have continually struggled over the designation of a 6 min. walk test as the measure of drug efficacy since it does not appear to correlate with any other clinical feature of pulmonary hypertension. You will see that in the future it will no longer be used as a primary endpoint in randomized clinical trials.	For KQ 3, we reported in Tables 23 and 24 the type of PAH that was studied (as reported in each published study). There were insufficient data to conduct patient-level analyses by different forms of pulmonary hypertension. We recognize (and discuss in the report) that 6MWD has limitations as an outcome measure, but this was the most commonly and consistently reported outcome across studies, and as such, we believe it should be included as an outcome in this systematic review.
TEP Member 3	General	While I was asked to be a technical advisor, my input occurred only after the project was already designed and set in motion. Another major problem with their evaluation of treatments is that the randomized trials for pulmonary hypertension have been totally dominated by the pharmaceutical industry. The 6 min. walk has been chosen as the primary endpoint because the regulatory authorities allow it, and the trials offer very little information beyond that. The trials by design are short in duration, limited in scope, and typically provide little information about the disease or how the patient benefits. The undustry has refused all requests to include the assessments of other tools, such as imaging or other biomarkers.	We augmented the discussion of the 6MWD, describing new information about validation and minimum important difference, and we discussed the limitations due to short duration and the paucity of published evidence on clinical significance of several of the clinical outcomes assessed. In the Discussion section called “Limitations of the Evidence Base,” we added a paragraph related to industry funding: “Although we did not find evidence for publication bias in a funnel plot of 6MWD outcomes, this does not ensure the absence of selective reporting. Modest but statistically significant effects seen in extant studies might nevertheless result from biases in study design or selective reporting of results. The extent to which the funding source may be related to this is unclear from our data; a majority of treatment trials (68%) were industry funded.”

Commentator & Affiliation	Section	Comment	Response
TEP Member 3	General	This initiative may be very helpful if it is redone 10 years from now, but I believe the major message is that the academic community, by relying on the pharmaceutical industry to direct the majority of clinical studies of this disease, has failed to identify clinical biomarkers of disease severity, and has failed to provide patients with adequate effective therapy to reduce mortality.	Thank you for your comment.
TEP Member 4	General	The report is clinically meaningful. The target population and audience are explicitly defined and the key questions are appropriate.	Thank you for your comment.

Commentator & Affiliation	Section	Comment	Response
TEP Member 7	General	The target populations are well defined, but the key questions should be revised as follows: KQ1 is problematic in that it addresses what should be considered 2 distinct patient groups: 1) Patients who are suspected of have PH due to their symptoms and clinical presentation 2) Asymptomatic patients at high risk of developing PAH. The former group should be used to address proper approach to diagnosis. The latter should be used to address screening. The question aims to determine the efficacy and safety of echocardiography and biomarkers in "screening" for both populations. In group 1, these tests are used to look for a diagnosis to explain the patient's symptoms. Here the strength of the tests is based on their specificity or positive predictive value. In group 2, the same tests are used to screen asymptomatic patients. In this population, the sensitivity and negative predictive value of these tests determine their usefulness as screening tools. Thus KQ1 would be better split into 2 questions that address the utility of echo and biomarkers in the diagnostic work up of dyspnea and the safety and efficacy of these tests as screening tools in asymptomatic patients at risk of PAH. This is particularly important, because it is not known if treatment is beneficial in patients with asymptomatic PH. That is, if an asymptomatic patient at high risk of PAH has a positive screening test, it is unclear if they should be taken to right heart catheterization and treated if PH is confirmed, or followed until they become symptomatic. Another problem is the use of the term screening, which usually refers to asymptomatic patients. On Page 19 is a section entitled "Echocardiography Plus Biomarkers for Screening PAH". However the study discussed in this section examined patients referred for PH evaluation who had symptoms of PH and were being evaluated for the symptom	We agree with the described framework. However, candidate studies often tested mixed populations. The type of screening we are interested in is sometimes called case finding. Aimed at populations that are at high-risk, one would still expect them to be free from signs or symptoms of the target condition. Many studies of high risk populations did not exclude patients with signs or symptoms, lumped together asymptomatic with symptomatic individuals or tested populations whose symptoms were attributed to a disease other than the target condition. We believe that although these studies do not conform to a strict definition of screening, they nevertheless provide information that is applicable to this situation.

Commentator & Affiliation	Section	Comment	Response
TEP Member 7	General	KQ2 Management of PAH: This key question would be better stated as "Use of noninvasive measures to monitor response to treatment" The clinical meaningfulness of the report is not robust. The questions put forth have high clinical relevance. The need for a method, test or evaluation algorithm to help determine who should receive right heart catheterization and who can be followed or dismissed is extremely important, especially as more patients are referred based on nonspecific test results. Also important is the need for a practical approach to monitoring response to therapy and determining when therapy should be altered or abandoned altogether in favor of lung transplantation. Long-term effectiveness and potential benefits of combination therapy need to be determined as more therapies continue to be improved for this disease. Unfortunately, the greatest finding from this report is that there is insufficient evidence to address these questions. Furthermore, some of the findings imply that noninvasive testing may be adequate for diagnosis and monitoring of PH and these results may serve to foster a false sense of security in caring for these challenging patients. Most experienced practitioners of PH, including this reviewer have discovered the frustration of not being able to consistently determine disease etiology, severity or response to treatment without invasive hemodynamic monitoring and the report needs to emphasize the limitations of non-invasive testing.	We agree that these are important clinical questions and that there are critical limitations in the published evidence.
TEP Member 8	General	The authors have done an amazing job in reviewing this literature of noninvasive diagnosis of pulmonary hypertension. There are a few things to add or phrase a bit differently but this report will be must reading for investigators who wish to study this field.	Thank you for your comment.
Public Reviewer - ACCP 1	General	The report reads like an assembly of facts, with very little synthesis of the data. I don't believe that this will be helpful to the pulmonary hypertension community in its current format. I find it confusing at times, with seemingly contradictory comments regarding the utility and accuracy of echocardiography, for example.	Thank you for your comment.
Public Reviewer - ACCP 1	General	The report seems to compare various assessments versus one another, when the reality is that providers tend to look at the larger picture created by looking at these assessments in combination with one another. In other words, what do the echocardiographic findings and BNP levels mean in the context of the 6MWD, FC, and hemodynamic measurements?	We have added a discussion of the recently published REVEAL risk score, which includes a broad spectrum of clinical information.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer - ACCP 1	General	The section on Pharmacotherapy (addressing KQ) seems to be particularly lacking in synthesis of the data, and this is perhaps the most important area being addressed.	Thank you for your comment.
Public Reviewer - ACCP 1	General	It seems as though our statement might need to begin by referring to the complete ACCP guidelines document published in 2004, and the update to the treatment portion of those guidelines published in 2007.	Thank you for your comment.
Public Reviewer - ACCP 1	General	In a general way, the review seems somewhat disproportionately focused on echocardiography.	Thank you for your comment.
Public Reviewer - ACCP 1	General	Comments about the utility of echocardiography appear to be somewhat inconsistent through the report.	Thank you for your comment.
Public Reviewer - ACCP 3	General	The review is excellent. The questions are important, but they also limit the analysis and the conclusion(s). One specific limitation is that the assessment of the prognostic value of biomarkers and echocardiography was limited to these two modalities. Therefore the work of the REVEAL Registry investigators to develop a multimodality prognostic score which includes echocardiography (parenthetically supporting the importance of pericardial effusion) and biomarkers was not considered or discussed. In my view, the REVEAL Score is the most highly developed and valid marker of prognosis for group 1 PAH patients currently available. An AHRQ review should include an analysis of the REVEAL Score.	We have added a discussion of the recently published REVEAL risk score, which includes a broad spectrum of clinical information.
Public Reviewer - ACCP 5	General	The document references the Venice classification rather than the Dana Point classification in multiple areas.	We updated our reference to Dana Point 2008, described in Simonneau G, Robbins IM, Beghetti M, et al. Updated clinical classification of pulmonary hypertension. J Am Coll Cardiol 2009; 54: S43-54.
Public Reviewer - ACCP 5	General	The term “screening” is applied at times to individuals “with symptoms that raise the suspicion of PAH” (see page ES-2, for example.) There is some elasticity to the definition of screening but in my book I’m not sure they are using the term optimally in that it is generally taken to imply asymptomatic patients.	The type of screening we are interested in is sometimes called case finding. Aimed at populations that are at high-risk, one would still expect them to be free from signs or symptoms of the target condition. Many studies of high risk populations did not exclude patients with signs or symptoms, lumped together asymptomatic with symptomatic individuals or tested populations whose symptoms were attributed to a disease other than the target condition. We believe that although these studies do not conform to a strict definition of screening, they nevertheless provide information that is applicable to this situation.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer - ACCP 5	General	Also relevant to screening is the fact that "relatively few data exist on the efficacy of treating PAH early in the disease course (WHO I-II.) Given this and the point above, I think we have to be very cautious about making recommendations that can be interpreted as favoring screening of asymptomatic, well functional patients unless they are in well-defined and well-studied high risk groups.	We agree and the key points note that there are insufficient data with regard to screening asymptomatic persons.
Public Reviewer - ACCP 5	General	They seem very enamored of the Bonderman ERJ study on BNP + echo, but I maintain some clinical skepticism and I would prefer we didn't overstate the evidence in favor of the approach that the study took.	We agree that this study suffers from some potential bias in the way the cohort was assembled and the way the reference standard was applied. Furthermore, it is so small that numerical estimates are imprecise. However, it does remain one of the only studies that gave information on the joint distribution of echocardiography and BNP in diagnosing precapillary PH, and given the focus of our key question, it has to be featured rather prominently. It seems fair to call for replication before widespread adoption of this strategy as there is a lot of uncertainty due to small numbers of subjects and confounding with ECG evidence or RVH.
Public Reviewer - ACCP 6	General	Briefly, this project was poorly thought out from the outset. The biggest mistake was lumping all PAH together, since most of us believe that vasoreactive PPH, IPAH, scleroderma PAH and congenital heart disease PAH are 4 distinct diseases with different presentations, pathobiology, response to therapy, and natural history. Thus this project was doomed to fail. The best I can say was that it confirmed what we already knew. However,, it taught us nothing.	Thank you for your comment. We believe that, despite the lack of strong conclusions, the project could motivate better research, which is one of the primary purposes of any systematic review.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer - ACCP 6	General	This project had as one of its initial goals an evaluation of the validity reliability and feasibility of echocardiography and biomarker testing for the screening diagnosis and management of PAH. Secondly they sought to determine whether the use of echocardiography or biomarkers would affect clinical decision-making and clinical outcomes. Finally they sought to determine which medications are effective for treating PAH and whether combination therapy is more effective than monotherapy. While AHRQ methodology has a proven track record for conducting exhaustive and accurate assessments of published data referable to diseases, I fear that this project was futile from the outset. It is unclear to me why the goals of the project were chosen, but in my opinion had they consulted the leading experts in the field this project would not have been done. The single largest flaw is the assumption that all of the different forms of pulmonary hypertension listed under the category of PAH are similar. The reality is they are very different. Secondly, the pulmonary hypertension community and regulatory authorities have continually struggled over the designation of a 6 min. walk test as the measure of drug efficacy since it does not appear to correlate with any other clinical feature of pulmonary hypertension. You will see that in the future it will no longer be used as a primary endpoint in randomized clinical trials.	For KQ 3, we reported in Tables 23 and 24 the type of PAH that was studied (as reported in each published study). There was insufficient data to conduct patient-level analyses by different forms of pulmonary hypertension. We recognize (and discuss in the report) that 6MWD has limitations as an outcome measure, but this was the most commonly and consistently reported outcome across studies, and as such, we believe it should be included as an outcome in this systematic review.
Public Reviewer - ACCP 6	General	While I was asked to be a technical advisor, my input occurred only after the project was already designed and set in motion. Another major problem with their evaluation of treatments is that the randomized trials for pulmonary hypertension have been totally dominated by the pharmaceutical industry. The 6 min. walk has been chosen as the primary endpoint because the regulatory authorities allow it, and the trials offer very little information beyond that. The trials by design are short in duration, limited in scope, and typically provide little information about the disease or how the patient benefits. The undustry has refused all requests to include the assessments of other tools, sich as imaging or other biomarkers.	We augmented the discussion of the 6MWD, describing new information about validation and minimum important difference, and we discussed the limitations due to short duration and the paucity of published evidence on clinical significance of several of the clinical outcomes assessed. In the Discussion section called “Limitations of the Evidence Base,” we added a paragraph related to industry funding: “Although we did not find evidence for publication bias in a funnel plot of 6MWD outcomes, this does not ensure the absence of selective reporting. Modest but statistically significant effects seen in extant studies might nevertheless result from biases in study design or selective reporting of results. The extent to which the funding source may be related to this is unclear from our data; a majority of treatment trials (68%) were industry funded.”

Commentator & Affiliation	Section	Comment	Response
Public Reviewer - ACCP 6	General	This initiative may be very helpful if it is redone 10 years from now, but I believe the major message is that the academic community, by relying on the pharmaceutical industry to direct the majority of clinical studies of this disease, has failed to identify clinical biomarkers of disease severity, and has failed to provide patients with adequate effective therapy to reduce mortality.	Thank you for your comment.
Public Reviewer - ACCP 6	General	This report follows the characteristic of other AHRQ reviews. It is very well organized, easy to read and well structured. It is thorough in its approach and provides a comprehensive overview addressing the questions posed.	Thank you for your comment.
Public Reviewer - ACCP 9	General	While key questions #1 and 2 address issues of interest and importance, and the findings perhaps valuable to inform certain clinical investigators of what has and has not been established, the lack of available data upon which to draw firm conclusions is neither surprising nor helpful to clinicians.	Thank you for your comment.
Public Reviewer - ACCP 9	General	The approach taken to KQ1 and KQ2, in which seemingly innumerable factors were assessed for correlations with others, is dizzying to read and not terribly informative. For KQ1, while goal of assessing whether a non-invasive test might be appropriate for diagnosis of PAH is indeed important, the approach taken seems to naively assume that a single parameter on echo might, for example, replace a RHC. A clinically more informed approach might have been to ask if combinations of echocardiographic findings might be useful in this manner, Although such an approach would almost certainly have resulted in similarly disappointing results, it would have at least made sense clinically. For KQ2, the approach seems taken seems to seek to answer whether any one non-invasive parameter is useful on its own as a means of monitoring patients. A more clinically useful approach would have been to assess whether adding one test to another is useful (e.g, what is the value of an BNP if one already has a dilated RV, or vice versa). This would have been not only clinically more important, but also more appropriate in an era when we are finally recognizing the need to assess value.	Indeed, we were interested in the diagnostic and prognostic value of combination of findings from echo, biomarkers and other clinical data. However, the vast majority of studies take a reductionist approach of looking at associations of one test or finding in isolation. We have added data from the REVEAL registry risk score which includes a broad spectrum of clinical information.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer - ACCP 9	General	The writing is mechanical and gives the reader the impression that the authors did not seek clinically useful illumination, but rather the accomplishment of a long list of isolated mathematical tasks. This is a shame, given the large and commendable effort undertaken in the assembly of a very large amount of information. It is only in the Discussion that one finally finds *synthesis* of the information (rather than mere assembly into tables and listed 'facts' in the text). Unfortunately, most of the points or conclusions reached in the discussion were evident to nearly anyone in this field ahead of time. There is little new that is now known.	We are attempting to follow the style guidance for comparative effectiveness reviews and clearly distinguish between the available data (in results) and our interpretation. Synthesis involves two considerations: (1) combining data statistically (meta-analysis), which we describe in the results sections, and (2) critical appraisal and interpretation, which falls more in the discussion. While nearly anyone in the field may not be surprised by any of the conclusions, there may be value to those both within and outside the field in a report that collates and describes these data, with some validation in reaching similar findings.
Public Reviewer - ACCP 9	General	The above is not at all meant to suggest that a rigorous review of a field that concludes we know less than we thought we did, or that the evidence base for certain things is not strong, is not of value. Such conclusions are often the most important contributions of systematic reviews and/or meta-analyses. But, to be consequential one would logically expect that such conclusions were not already the current consensus among most experts within a field.	Thank you for your comment.
Public Reviewer - ACCP 9	General	It is alarming to see the term "screening" used anywhere in a document from AHRQ refer to a test performed for evaluation of a symptom. Screening, by definition, involves looking for disease in an asymptomatic individual.	The type of screening we are interested in is sometimes called case finding. Aimed at populations that are at high-risk, one would still expect them to be free from signs or symptoms of the target condition. Many studies of high risk populations did not exclude patients with signs or symptoms, lumped together asymptomatic with symptomatic individuals or tested populations whose symptoms were attributed to a disease other than the target condition. We believe that although these studies do not conform to a strict definition of screening, they nevertheless provide information that is applicable to this situation.
Public Reviewer - ACCP 9	General	Tadalafil does not appear to have been included in the results of KQ3. It needs to be. It does appear to have been in the search strategy.	We identified 3 RCTs that evaluated the efficacy of tadalafil (Table 23). We have included the findings of these 3 studies in the meta-analyses for the outcomes reported in these studies.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer - ACCP 9	General	My apologies if I missed it, but why has there been no formal assessment of publication bias, particularly regarding KQ3? This seems essential in a field so clearly dominated by pharmaceutical industry-sponsored studies.	In the revised report, we assessed possible publication bias by creating and interpreting a funnel plot of all the studies that reported 6MWD, and we identified abstracts and registered completed trials that remain unpublished at the beginning of the Results section. We note in the Results that the vast majority of the studies included in the KQ 3 analyses were industry-funded trials.
Public Reviewer - ACCP 9	General	The authors have used an outdated version of the Who classification for PH	We updated our reference to Dana Point 2008, described in Simonneau G, Robbins IM, Beghetti M, et al. Updated clinical classification of pulmonary hypertension. J Am Coll Cardiol 2009; 54: S43-54.
Public Reviewer - ACCP 9	General	The reader is not provided with details to help understand what led to the downgrading of a given study's quality rating (e.g., from good to fair). Such information is important as while certain criteria may result in similar downgrading, they may have differing importance according to the question asked and study design.	As described in the Methods section, individual study quality ratings were evaluated based on features described in Table 3 and grouped in categories good, fair or poor. Specific ratings for each study are provided in Appendix D along with features that limit the applicability of the findings. (Note that applicability concerns were not given a single rating as study quality was. We made an editorial decision not to present the myriad data elements that went into the quality grade due to space constraints. Quality ratings were reached as the consensus of two (and sometimes three) reviewers of the article; disagreements were resolved by consensus.
Public Reviewer - ACCP 9	General	The exclusion of studies prior to 1995 in a field with few published studies (e.g. RCTs for epoprostenol) seems foolish unless there is a very good reason to do so. The reason stated in the methods section seems to relate only to KQ1 and KQ2, not to KQ3, so the authors need to reconsider this. As KQ3 seems to be the only portion of this SR/MA that might inform clinically actionable recommendations, it would seem all the more important to include all meaningful data unless there are very strong reasons why the study designs/quality are problematic to a degree not found in those studies since 1995 that have been included (e.g. Rubin et al. 1990, PMID 2107780; Barst et al 1994, PMIDL 8053614; Rich et al PMID 1603139)	The 1995 cutoff was chosen based on the FDA approval of Flolan in 1996 as the first of the newer vasodilator treatment for PAH. Unfortunately, this was not early enough to capture the first trial by Rubin et al., 1990. As a result of reviewer's comments, we have revised the searches for KQ 3 to include 1990–1994 and screened the resulting citations. Consistent with the reviewer's suggestions, the only additional study that meets our inclusion criteria is Rubin. We amended the protocol to reflect this change. The study by Rich et al. was ineligible for inclusion because it did not include a comparator group. The study by Barst et al. was included as extension study of the RCT by Rubin et al., 1990.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer - ACCP 10	General	Before addressing comments of each of the key questions posed by AHRQ, I want to express that while well intentioned, I do not think that the Key Questions chosen by AHRQ were necessarily the correct ones to inform our guidelines recommendations. As reflected in our joint discussion at ATS, it was almost a foregone conclusion that Key Questions #1 and #2 would not have sufficient evidence to allow a definitive conclusion to be shared in any guideline format. Indeed, it seems that a great deal of effort and rigor went into coming up with a conclusion that our PAH expert panel already knew. In reading the draft, I got the distinct sense that those composing it were not really content experts in PAH. It read rather mechanically and had a great deal of redundancy. There were also numerous typographical errors (though I am sure our colleagues at Actelion are relieved to know their drug had not effect on patient morality…).	The Key Questions are not chosen by AHRQ, but rather are developed as a result of discussions held with Key Informants. Also, the Key Questions are posted for public comment during the topic refinement stage and are further refined with input from the Technical Expert Panel during the writing of the report.
Public Reviewer - ACCP 10	General	It is my suggestion that both Key Questions # 1 and 2 be disregarded and not used to formulate any guideline suggestions in general as they do not address clinically important issues. Furthermore there is not adequate quality data to draw from regarding these questions as I think the AHRQ realized in the end.	Thank you for your comment.
Public Reviewer - ACCP 10	General	Regarding the KQ1 and 2 more detailed discussion again I found no surprises and their findings were as expected frankly. However, I found it odd that 2 studies of the accuracy of echo in PAH were omitted. Specifically Arcasoy S, et al, Am J Resp Crit Care Med, 2003 and Fisher MR, et al, Am J Resp Crit Care Med, 2009 were glaring omissions of 2 well done and important studies that I feel should have been included.	We added these studies to our analysis for KQ 1 in final report; both studies were identified by our literature search. Arcasoy et al. was excluded early in our screening process, and probably erroneously. Fisher et al. was excluded very late in the process of writing the report; however, some similar studies were included, so we included it as well.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer - ACCP 2	Do KQs address clinically important issues?	I feel the key questions do address clinically important issues, in particular KQ1 and KQ3. Since I am a rheumatologist, I feel that an analysis of the published literature regarding screening of high-risk populations (KQ1) is timely and relevant. As indicated by the comparative review performed, there are many studies of fair to good quality evaluating the utility of NT-proBNP and multiple echocardiographic parameters for screening of symptomatic and asymptomatic patients for PAH. This is extremely relevant to the SSc population of patients, and potentially to other connective tissue diseases. Data from the PHAROS study indicates that mortality in this cohort of patients who are routinely screened for PAH may be better than other cohorts of patients followed in the community (Chung et al. Arthritis Rheum 2011;63(10S):S673). However, guidelines for when to refer for right heart catheterization have not yet been solidified. In addition, the role of NT-proBNP in screening patients for PAH has not been defined. The review performed demonstrates that a normal NT-proBNP may be useful in ruling out PAH in patients with elevated sPAP on echo. This supports using this biomarker in this patient population, however, currently most insurance companies are not reimbursing this as a screening test in SSc patients. Publishing such a guideline would be helpful to standardize using this test as a routine, reimbursable screening test. In addition, reviewing the echocardiographic parameters that correlate best with RHC, and the cut-offs with highest sensitivity and specificity in high-risk populations, like patients with SSc, will guide clinicians in their referrals for RHC. This review verifies the correlation of sPAP on echo with hemodynamic parameters, and also reviews some of the more novel echo measures, such as TRV/VTI_{RVOT} that can be useful in screening patients for PAH.	Thank you for your comment.
Public Reviewer - ACCP 2	Do KQs address clinically important issues?	KQ2 verifies the use of BNP and echo in monitoring patients with PAH, with the caveat that the former only has moderate correlation with hemodynamics and exercise capacity. Pericardial effusion was the only parameter that had a significant predictive value for mortality. Although useful information, the results compiled for KQ2 do not suggest that a change in guidelines for patient monitoring is necessary.	Thank you for your comment.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer - ACCP 2	Do KQs address clinically important issues?	KQ3 evaluates the RCTs completed for approved PAH-specific therapies, both mono and combination therapies. It is definitely timely to review and compile the results of these studies given the multiple approved agents over the past couple of decades. The results are relevant to the clinical care of patients with PAH. Although there is still insufficient data to assess effects on mortality, and minimal data to assess effects on hemodynamics, the positive benefits from PDES-I and ERA on decreasing hospitalizations and improving exercise capacity validate guidelines to initiate these therapies in patients with FC II-III disease. The utility of these agents in asymptomatic patients who are at high risk for progressive disease (ie. High BNP, SSC) needs further study. There were few studies evaluating combination therapy, and the data supporting this had a low strength of evidence. It may be that further studies with less heterogeneity need to be performed before a good assessment of combination therapy can be performed. The results related to adverse effects are not unexpected, but are helpful in counseling patients regarding potential side effects.	Thank you for your comment.
Public Reviewer - ACCP 3	Do KQs address clinically important issues?	KQ3 is most important. KQ2 is moderately important (and gaining important in an era of cost control/consciousness). KQ1 may be the least important now, but is certainly important for the future.	Thank you for your comment.
Public Reviewer - ACCP 4	Do KQs address clinically important issues?	KQ1L Certainly an important clinical issue	Thank you for your comment.
Public Reviewer - ACCP 4	Do KQs address clinically important issues?	KQ1L Need for a method, test or evaluation algorithm to help determine who should receive right heart cath and who can be followed or dismissed is huge especially as more patients are referred based on echo findings alone	Thank you for your comment.
Public Reviewer - ACCP 4	Do KQs address clinically important issues?	KQ2: This key question would be better stated as “Use of noninvasive measures to monitor response to treatment”	Thank you for your comment.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer - ACCP 4	Do KQs address clinically important issues?	KQ2: Also agree this is highly relevant issue.	Thank you for your comment.
Public Reviewer - ACCP 4	Do KQs address clinically important issues?	KQ2: Present guidelines suggest following a variety of clinical, laboratory and echo findings but put the final decision on what represents improvement, lack of change or deterioration on the back of the clinician. Identification of objective variables that could be used to ascertain treatment response or lack there of should enable patients to have treatment increased more rapidly and decrease the high number of patients who die from PAH without receiving prostanoid therapy.	Thank you for your comment.
Public Reviewer - ACCP 4	Do KQs address clinically important issues?	KQ3: Obvious clinical importance	Thank you for your comment.
Public Reviewer - ACCP 4	Do KQs address clinically important issues?	KQ3: Agree with stated need for head to head comparator studies between different therapies, monotherapy versus combined therapy, and long-term effectiveness.	Thank you for your comment.
Public Reviewer - ACCP 5	Do KQs address clinically important issues?	Yes, the key questions address clinically important issues. By and large, the right outcomes have been evaluated for each question and appropriate evidence seems to have been included.	Thank you for your comment.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer - ACCP 7	Do KQs address clinically important issues?	These are important questions that address both screening/management, and cost-effectiveness issues for a rare disease. Questions 1 and 2 address issues that are raised frequently in clinical practice and clinical trial design when selecting appropriate endpoints. For example, practitioners have become less inclined to perform cardiac catheterizations and rely more heavily on echo and BNP. While the echo seems to correlate with the RHC for estimation of sPAP, it still leaves the practitioner short on other valuable data that are critical for the management of PAH patients, including CI, PCWP and RAP. And with the current push to rely more heavily on less invasive measures, the role of Echo and biomarkers has become very important. The question of screening for asymptomatic patients remains unanswered in this review but is a very important question for the group particularly in preparation for guidelines.	Thank you for your comment.
Public Reviewer - ACCP 7	Do KQs address clinically important issues?	With regard to question #3, the key question of how agents compare to one another remain unanswered due to heterogeneity of studies and limited follow up. However, the use of combination vs. placebo and vs. monotherapy does appear more favorable with respect to reduction in hospitalizations which will be an important topic for us to cover. Further, the finding that hemodynamic improvement occurred with monotherapy vs. combination may help with clinical trial design and endpoint development.	Thank you for your comment.
Public Reviewer - ACCP 8	Do KQs address clinically important issues?	A very detailed document but does not quite address the questions that are clinically pertinent.	Thank you for your comment.
Public Reviewer - ACCP 2	Are right outcomes evaluated for KQs, any outcomes missing?	For KQ2, possible outcomes to assess in fuller detail would be escalation to prostanoid therapy; escalation to combination therapy; transplantation or atrial septostomy.	We focused on those outcomes that were most widely reported in the literature.
Public Reviewer - ACCP 2	Are right outcomes evaluated for KQs, any outcomes missing?	For KQ3, a review of the effect of immunosuppression as adjunctive (with PAH-specific therapies) or first-line therapy in the various CTD-APAH would be helpful and directly impact the care of patients with CTDs.	We appreciate the suggestion; however, it is beyond the scope of this project.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer - ACCP 2	Are right outcomes evaluated for KQs, any outcomes missing?	I suggest including the following two articles to address KQ1: Shah AA, Chung SE, Wigley FM, et al: Changes in estimated right ventricular systolic pressure predict mortality and pulmonary hypertension in a cohort of scleroderma patients. Ann Rheum Dis 2012 Aug 11. This is a recent study supporting the utility of serial echocardiograms in the screening of SSc patients for PAh and should be included in the review. Thakkar V, Stevens WM, Prior D, et al: N-terminal pro-brain natriuretic peptide in a novel screening algorithm for pulmonary arterial hypertension in systemic sclerosis: a case-control study. Arthritis Res Ther 2012 Jun 12;14(3):R143. This is a recent publication evaluating NT-proBNP in a screening algorithm for PAH in SSc patients.	Thank you for calling these studies to our attention. We added these studies to our analysis in the final report.
Public Reviewer - ACCP 3	Are right outcomes evaluated for KQs, any outcomes missing?	For the most part the right outcomes are evaluated. Although it is not an outcome, I do think the multimodality REVEAL risk calculator should be included with echo and biomarkers for KQ2. It incorporates echo and biomarker.	We have added a discussion of the recently published REVEAL risk score, which includes a broad spectrum of clinical information.
Public Reviewer - ACCP 4	Are right outcomes evaluated for KQs, any outcomes missing?	KQ1L Evaluation of BNP levels is appropriate as this is the only biomarker that has been used enough to provide reasonable data.	Thank you for your comment.
Public Reviewer - ACCP 4	Are right outcomes evaluated for KQs, any outcomes missing?	KQ1: Primary outcome for these studies in presence or development of PAH. Other useful outcomes would include examination of ability of these biomarkers to identify patients with worse functional capacity or survival	The association with functional capacity and survival was addressed as part of KQ 2
Public Reviewer - ACCP 4	Are right outcomes evaluated for KQs, any outcomes missing?	KQ1: Most of the conclusions have been drawn for studies using NT-BNP, a test that is not readily available to most practitioners.	We focused on the outcomes that were most widely reported; it is interesting to note that this test is so predominant in the literature, yet not widely available.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer - ACCP 4	Are right outcomes evaluated for KQs, any outcomes missing?	KQ1: Most of the BNP studies evaluated patients with PAH associated with SSc raising the question of whether same results can be expected with other forms of PAH or populations at high risk of PAH	Thank you for your comment.
Public Reviewer - ACCP 4	Are right outcomes evaluated for KQs, any outcomes missing?	KQ1: Most outcome measures for echocardiography are good. Additional outcome measures that should be examined include: CO on RHC correlation with TAPSE and FAC RVEDP on RHC correlation with TAPSE and FAC RAP on RHC correlation with RAP on echo	Few data were available on these analyses. We did look carefully at the question of RAP estimation because of its use as one component in estimating sPAP.
Public Reviewer - ACCP 4	Are right outcomes evaluated for KQs, any outcomes missing?	KQ1: The finding that TRV/VT_{RVOT} more consistently correlated with RHC than TRV/TG/sPAP may be noteworthy.	Thank you for your comment.
Public Reviewer - ACCP 4	Are right outcomes evaluated for KQs, any outcomes missing?	KQ1L For echocardiography, screening populations were SSc, liver transplant and Sickle cell disease. For liver transplants, diagnostic criteria use PVR>120 for several studies which may be too low. This cut off my improve sensitivity, but decrease specificity.	Thank you for your comment.
Public Reviewer - ACCP 4	Are right outcomes evaluated for KQs, any outcomes missing?	KQ1: For other studies, reference criteria ranged from 2-8 woods units making comparisons difficult.	Thank you for your comment.
Public Reviewer - ACCP 4	Are right outcomes evaluated for KQs, any outcomes missing?	KQ2: Outcome measure in table 16(Page 64) are appropriate	Thank you for your comment.
Public Reviewer - ACCP 4	Are right outcomes evaluated for KQs, any outcomes missing?	KQ2: Poor predictive value of BNP and echocardiographic findings appear contrary to results of some individual studies	Thank you for your comment.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer - ACCP 4	Are right outcomes evaluated for KQs, any outcomes missing?	KQ3: Use of the 6MWD as outcome for pharmacotherapy is appropriate considering its extensive use in clinical trials and clinical practice	Thank you for your comment.
Public Reviewer - ACCP 4	Are right outcomes evaluated for KQs, any outcomes missing?	KQ3: Hemodynamic outcomes would be attractive but apparently insufficient data exists	Thank you for your comment.
Public Reviewer - ACCP 4	Are right outcomes evaluated for KQs, any outcomes missing?	KQ3: It is unclear to me why hospital admission rate is used instead of the composite index of time to clinical worsening	Composite endpoints are problematic to assess if individual endpoints making up the composite are not described. Composite endpoints, in essence, assign equal importance to different events in the composite. Furthermore, they are hard to compare when defined differently among studies. We assessed mortality and hospitalization separately; however, outcomes such as transplantation were even rarer than death and could not be examined separately. See Ferriera-Gonzalez I, Busse JW, Heels-Ansdell D, et al. Problems with use of composite end points in cardiovascular trials: systematic review of randomized controlled trials. BMJ 2007 doi:10.1136/bmj.39136.682083.AE (published 2 April 2007).
Public Reviewer - ACCP 7	Are right outcomes evaluated for KQs, any outcomes missing?	There was a lack of pediatric data included in this analysis and groups were very heterogeneous.	Thank you for your comment.
Public Reviewer - ACCP 7	Are right outcomes evaluated for KQs, any outcomes missing?	There was a paucity of data on how to screen asymptomatic at risk groups which we will need to address further.	We discuss this limitation in the Applicability section.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer - ACCP 7	Are right outcomes evaluated for KQs, any outcomes missing?	Additional biomarkers or combinations of echo measures should be evaluated.	We sought data for a large number of biomarkers and then selected them based on availability of data.
Public Reviewer - ACCP 8	Are right outcomes evaluated for KQs, any outcomes missing?	Is it possible to compare different classes of agents?	Class comparisons were performed in our analyses.
Public Reviewer - ACCP 8	Are right outcomes evaluated for KQs, any outcomes missing?	Does a subpopulation respond better to a certain agent?	We sought data on subgroups in our review and reported where available.
Public Reviewer - ACCP 8	Are right outcomes evaluated for KQs, any outcomes missing?	Response to a particular agent based on functional class	We presented a sensitivity analysis by functional class in the discussion section; there were too few studies of particular agents to perform a similar analysis by agent.
Public Reviewer - ACCP 4	Appropriate evidence included? Important studies missing?	KQ1: Most of the conclusions have been drawn for studies using NT-BNP, a test that is not readily available to most practitioners	We focused on the outcomes that were most widely reported; it is interesting to note that this test is so predominant in the literature, yet not widely available.
Public Reviewer - ACCP 4	Appropriate evidence included? Important studies missing?	KQ1: Most of the BNP studies evaluated patients with PAH associated with SSc raising the question of whether same results can be expected with other forms of PAH or populations at high risk of PAH	Thank you for your comment.
Public Reviewer - ACCP 4	Appropriate evidence included? Important studies missing?	KQ2: These studies rely heavily on SSc and liver disease. Few studies of IPAH. I have some concerns about the inclusion of Sickle cell disease because of the multi-faceted causes of PAP elevation in this patient group	Thank you for your comment.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer - ACCP 4	Appropriate evidence included? Important studies missing?	KQ2: Is it possible to include Steve Mathai's recently published article on MID of 6MWD?	We have added a supporting citation for an emerging consensus regarding MID for 6MWT of 33 meters (Mathai et al 2012).
Public Reviewer - ACCP 4	Appropriate evidence included? Important studies missing?	KQ3: All the major clinical trials appear to be included.	Thank you for your comment.