Comparative Effectiveness Research Review Disposition of Comments Report

Research Review Title: *Comparative Effectiveness of Diagnosis and Treatment of Obstructive Sleep Apnea in Adults*

Draft review available for public comment from October 27, 2010 to November 24, 2010.

Research Review Citation: Balk EM, Moorthy D, Obadan NO, Patel K, Ip S, Chung M, Bannuru RR, Kitsios GD, Sen S, Iovin RC, Gaylor JM, D'Ambrosio C, Lau J. Diagnosis and Treatment of Obstructive Sleep Apnea in Adults. Comparative Effectiveness Review No. 32. (Prepared by Tufts Evidence-based Practice Center under Contract No. 290-2007-100551). AHRQ Publication No. 11-EHC052-EF. Rockville, MD: Agency for Healthcare Research and Quality. July 2011. Available at: www.effectivehealthcare.ahrq.gov/reports/final.cfm.

Comments to Research Review

The Effective Health Care (EHC) Program encourages the public to participate in the development of its research projects. Each comparative effectiveness research review is posted to the EHC Program Web site in draft form for public comment for a 4-week period. Comments can be submitted via the EHC Program Web site, mail or E-mail. 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: www.effectivehealthcare.ahrq.gov
Published Online: October 25, 2011

Reviewer[1]	Section[2]	Reviewer Comments	Author Response[3]
1	Cover page	Even though the draft report has been redacted to conceal the authors and which EPC conducted the report, the cover page in the downloaded file lists the authors names and affiliations defeating the purpose of the redaction.	Apologies. We attempted to redact this several times.
1	P15 ln 12	use of RE meta-analysis for as few as 3 studies creates important limitations in the usefulness of the results – note that rather than describing the best estimate of a true effect (under the fixed effects) the RE results in an estimate of a mean effects of studies with acknowledged disparate effects. It has little meaning clinically. It seems meta-analyses were often performed in the presence of significant heterogeneity, acknowledging clinical and methodologic heterogeneity. I'm not sure those analyses add much to the forest plots.	While we don't disagree, the decisions of whether, when, and how to meta-analyze are judgment calls. We decided to set a standard approach for consistency and to use the conservative approach of using the random effects model.
1	P15 ln 40	Exec summary KQ1 – describe Type III and IV monitors before these are discussed/defined. My understanding is the exec summary should standalone. Furthermore, this is not defined in Table 1 Acronyms and Abbreviations. The definition does not appear until a footnote to Table A on page xxiv.	We have copied over the fuller definitions of the different types of monitors from our 2007 report into the Introduction of this report. We have given a succinct, but better, definition of Types III and IV monitors just before the description of the results for these monitors. It is not an acronym or abbreviation, so is not listed in that table. Page xv, 2nd to last paragraph. Pages 5&6
1	P16 ln 47	Key question 4 has key question 3 glommed on to the end.	Thank you. We have corrected this.

Reviewer[1]	Section[2]	Reviewer Comments	Author Response[3]
1	P17 ln 47-50	"Despite the lack of evidence for an effect of PAP on clinical outcomes, given the large magnitude of effect on the intermediate outcomes of AHI, ESS, and arousal index, the strength of evidence that PAP is an effective treatment for the relief of signs and symptoms of sleep apnea was rated moderate." In the Analytic Framework there is no key question regarding the link between intermediate outcomes (of AHI etc) and health and related outcomes (which include symptoms). In the absence of such data. If the magnitude of effect on intermediate outcomes is so large, and the link between intermediate outcomes and health and related outcomes is clinically important, then one should be able to detect an effect if it exists, with caveats for inadequate power and other methodological shortcomings. Given that 22 studies were available, but showed only weak evidence of an effect and were consistent with no difference between sham and true PAP, then I think you must draw the conclusion that the strength of evidence that PAP is an effective treatment for relief of signs and symptoms of sleep apnea is weak. I reviewed the definition of moderate in the methods section (p54 ln 3-13), and apparently you are using this term consistently with this definition; however, I would be interested in your justification for this decision given the design of the Analytic Framework, and the lack of observed effect on measured clinical outcomes.	The 22 studies found strong evidence of a large effect on AHI and ESS between PAP and sham PAP (similar for PAP vs no treatment). We could not detect an effect on objective clinical outcomes because of a lack of evidence. Studies have not evaluated objective clinical outcomes. We agree it was a judgment call, but the methodologists and the domain expert agreed that this was a reasonable conclusion. We provide an explanation for how we arrived at this strength of evidence, so the readers can decide for themselves if they agree with the approach. We have revised the text in the Methods section (pages 24-25) on the definitions of the strengths of evidence. They now are aligned more with the Methods guide. We have clarified the term "clinical outcomes" to clarify that we mean "objective clinical outcomes", eg, mortality or clinical events.

Reviewer[1]	Section[2]	Reviewer Comments	Author Response[3]
1	P18 ln 8-17	is there any data on the minimal clinically important difference for ESS or oxygen saturation? Knowing this would be helpful for interpreting the reported differences in means.	We did not systematically review this, but we found no true evidence of clinically important differences or changes. However, experts informed us that a change in ESS of 1 point is clinically significant (by consensus). We had no clear information about an oxygen saturation threshold. We have added these into our new descriptions of outcomes in the Methods chapter (Page 19).
1	P18 ln 24-25	"Largely due to small sample size, the studies mostly had imprecise estimates of the comparative effects" check grammar … the estimates are imprecise largely due to small sample size-> largely due to small sample size, the estimates …	Fixed.
1	P18 ln 35-36	"There is a low strength of evidence of no substantial difference in compliance or other outcomes between the proprietary device C-Flex and CPAP" Wording here suggests equivalence, but absence of evidence of difference is not the same as evidence of no substantial difference.	This has been reworded to state: "No statistically significant differences in compliance or other outcomes were found between C-Flex and CPAP. The strength of evidence for this finding is rated low because of the mixed quality (Bs and Cs) of the four primary studies." Page xviii, last paragraph Page 83, Summary
1	P 25 ln 43	"…found that those who had mandatory PSG had nonsignificantly shorter hospital stays" More properly, found that those who had mandatory PSG had shorter mean hospital stay, but the observed difference was not statistically significant. Is it fair to conclude that a finding of the study that mandatory PSG had shorter stays? I don't particularly like this wording because it arguably attributes too much importance to nonsignificant differences. I suppose this might be worth pointing out if the magnitude of the difference were clinically important, but that it failed to reach statistical significance because of low statistical power.z	We agree. We have toned down the language in the summary of the Results in Chapter 3 (page 49) and have altered the summaries in other places to state that no significant differences were found (pages xvi, and the Summary of findings tables in the executive summary and Chapter 4)
1	P29 ln 37 (Table A)	Abbreviation UPPP not defined in text at first use, although it does appear in Table 1 Acronyms and Abbreviations (which is not part of standalone Executive Summary).	Thank you. Corrected.

Source: www.effectivehealthcare.ahrq.gov
Published Online: October 25, 2011

Reviewer[1]	Section[2]	Reviewer Comments	Author Response[3]
1	P43 Ln 21	"The The…"	Thank you. We did not find this at this location, but we fixed it elsewhere.
1	P 53	Is there an empirical basis for the assumption of fixed correlation between baseline and final values within patient? And the specific value of 0.5?	No. This is an approach we have used for several years. You have, though, inspired us to pursue this question as a methods project.
2	Quality of the Report	Superior	Thank you.
2	General Comments	In general I think the reports is meaningful. Target audience appears adequate but I think it may need additional definitions (I found it difficult to find delineation of this and it seemed too general). The key questions are appropriate and well formulated	We have discussed this with our Task Order Officer. The standard Preface deals with the issue of audience and that, unlike translational products, CERs do not target specific audiences.
2	Introduction	I think it should clarify that age is also an important risk factor, with some data supporting that those older than 60 have a significantly higher prevalence than younger patients (age 20-29), this being also a common finding among practicing sleep physicians.	Our description of an association has been improved. We found different data than is stated here.
2	Methods	I agree with the selection criteria and everything else, but feel there should have also been some data regarding combined therapies for OSA in the search criteria, such as the use of oral devices in conjunction with CPAP.	The search criteria covered combination therapies. The combinations do not need to be explicitly included in the search.

Source: www.effectivehealthcare.ahrq.gov
Published Online: October 25, 2011

Reviewer[1]	Section[2]	Reviewer Comments	Author Response[3]
2	Results	The amount of material presented is adequate. The studies are well presented. The heterogeneity of some studies truly questions the validity of some of this data. Figures and tables are adequate. As noted below. I think the use of oral devices may be inadequate in patient's with severe apnea, and this should be highlighted. Also, the use of combined modalities of therapy was not investigated, such as oral devices in combination with CPAP or positional therapy in combination with CPAP. I see some mention of positional alarms but regarding forcing devices for positional therapy in the executive summary this is not clearly stated. This is frequently used, at least in those cases where studies suggest improvement and some times in combination with CPAP, and the studies mentioned show overall no improvement over CPAP, but I think this is still valid therapy in a subset of patients. I think there needs to be further elaboration in this regard, as this is also an area of further research and potential therapy in a subset of patients. Would also include this reference: J Clin Sleep Med. 2010 Jun 15;6(3):238-43. Comparison of positional therapy to CPAP in patients with positional obstructive sleep apnea. Permut I, Diaz-Abad M, Chatila W, Crocetti J, Gaughan JP, D'Alonzo GE, Krachman SL. Temple University School of Medicine, Philadelphia, PA 19140, USA. Would also include some data on treatment of nasal obstruction and its effect in OSA.	Combination therapies were included, though studies of them may not have met general eligibility criteria. We have made this more explicit in the Methods (page 17). Permut 2010 was found in the updated search performed during the review process. The study has been added To add clarity, we have added the following sentence to the Methods section of the Executive Summary (page xv) and of the full report (page 15, near bottom, Chapter 4 page 131): Of note, where interventions are not discussed (either diagnostic tests or treatments), this does not imply that the interventions were excluded from analysis (unless explicitly stated); instead, no studies of these interventions met eligibility criteria.

Reviewer[1]	Section[2]	Reviewer Comments	Author Response[3]
2	Discussion/ Conclusion	The major findings are clearly stated, but some, as noted above, need additional clarification. The limitations of the studies in general are well defined. Regarding the future research section, I think there needs to be more emphasis in identifying the patient groups that are likely to benefit from therapy on the long run. For instance; are asymptomatic (patients without hypersomnolence, etc) with mild to moderate OSA likely to have any benefit, even if they have associated HTN or a mood disorder not attributed to their OSA? Does this depend on the degree of severity of their OSA? Many patients with OSA also have associated central apneas, despite no clear other causes, and perhaps these being induced by the use of CPAP. Does this carry a clinical connotation that may negate the benefits of treating patients with OSA, if not severe and/or those without associated hypersomnolence or other comorbidities. These are important considerations for future research. I also think it is important to consider that individualized therapy is important. Hence, a study may not demonstrate that in general compliance changes with one mode of therapy over another, yet in certain cases, a mode of therapy may affect that individual's compliance. I am not sure how to demonstrate this, yet I feel additional research in selected groups is needed.	Thank you. We have added in bullets specific to the need for studies in different disease severity populations and in those with mixed disease (central and obstructive). We think we have made sufficient comment on future research of compliance.

Reviewer[1]	Section[2]	Reviewer Comments	Author Response[3]
2	Clarity and Usability	I think this is well organized and structured and information is clear, though it appears there is insufficient strong evidence to support clear clinical guidelines and hence policy. I think the study highlights the fact that current practice may not be ingrained in strong clinical evidence, and that perhaps current practice may be leading to some over treatment of individuals that may have no clinically significant benefit from this therapy, such as asymptomatic OSA patients like I mentioned above.	Thank you
3	Quality of the Report	Superior	Thank you
3	General Comments	This is an update of the 2007 AHRQ review of the topic. It is focused on the most critical and clinically relevant questions in the field. It is comprehensive in scope. It is thorough in its review. It is sound in the conclusions drawn.	Thank you
3	Introduction	The introduction provides the reader with a well-written and accurate overview of the issues to be addressed and the background to date pertaining to those issues. I am not sure that the statement that polysomnography is "poorly tolerated" (p9) is entirely justified. Most patients tolerate the procedure quite adequately. Descriptors such as "cumbersome," time-consuming" or "labor intensive" might be more apt.	We agree and have changed poorly tolerated to “often inconvenient” in the Abstract and the Introduction.
3	Methods	Methods are clearly described in sufficient detail. Inclusion and exclusion criteria are rigorous, though appropriate. The search strategies were generally well-defined though some readers may wish for more explicit description of specific search terms. Statistical analysis appears sound although this is not my area of expertise.	Thank you

Source: www.effectivehealthcare.ahrq.gov
Published Online: October 25, 2011

Reviewer[1]	Section[2]	Reviewer Comments	Author Response[3]
3	Results	The results sections, with appendices and extraction sheets is overwhelmingly thorough. Inevitably, some specific data that may be informative are not found in the summary tables; in some cases this reflects unavailability of such data (eg information re specific subtypes within both diagnostic and treatment data) - the lack of such data is made clear in the text. In other cases, certain information (eg specific information re methodology for auto-titration CPAP studies) is omitted, but this is not a serious deficiency. I am not aware of major investigations that have been omitted. Much is made of the poor compliance with PAP treatment and this is indeed a key issue. However, we dot have have clear information regarding how partial compliance (eg patients who do not meet the s/w arbitrary criteria of >4h for 70% of nts) may or may not yield some benefit. One must also acknowledge the reality that compliance with other treatment modalities/recommendations can be problematic as well. The issue of treatment effectiveness for PAP vs. MAD (or surgery) by subtype is especially important and, while such analysis is often not possible, any further effort to glean such information from existing data would be helpful.	We have removed this arbitrary definition of compliance. We have also made clearer throughout the results how compliance was measured

Reviewer[1]	Section[2]	Reviewer Comments	Author Response[3]
3	Discussion/ Conclusion	The conclusions drawn from the data presented are reasonable and follow logically from the stated parameters and available data. Not surprisingly, they are consistent with previous conclusions drawn from extant data in the 2007 report. The shortcomings of the data are made clear and, for the most part, the specific recommendations made for future research are appropriate. In this regard, the future recommendations do not identify analysis of subtypes (eg by severity) as a key component of future studies re diagnostic accuracy of various tools. This represents an important clinical issue which should be included.	We have added a future research recommendation for studies that evaluate the appropriate tests based on the type and severity of their symptoms.
3	Clarity and Usability	The report is quite well-organized and easy to follow. There is substantial redundancy in conclusions and summary through the report but this is not necessarily a bad thing and, to some extent, reflects the nature of the organization of the report. The conclusions and recommendations are highly relevant to the clinical enterprise and identify the most important issues for further research, with the caveats noted above. Unfortunately, the report leaves the field in much the same situation as it has been with respect to key issues of diagnosis, treatment modalities and outcome. This is the nature of the data but serves a useful purpose if only to remind researchers and clinicians of what gaps exist and what needs to be done to fill those significant gaps	We agree. No revisions made.

Reviewer[1]	Section[2]	Reviewer Comments	Author Response[3]
4	General	This is a well written review paper that examines a variety of important and clinically relevant questions concerning the diagnosis and treatment of sleep apnea. Numerous theoretical and research issues are also discussed that help the reader to carefully interpret the findings and their limitations. The review seems thorough, relevant data extracted, and the studies adequately summarized. Most of the figures are helpful and the tables informative. However, there are a few issues that are somewhat troublesome that should be addressed and some minor points that need clarification.	Thank you.
4	Introduction	The introduction is well-written and covers the various areas reviewed in a clear and concise manner. General information is included that allows the reader to get a good overview of the numerous issues involved in diagnosing sleep apnea and the impact of the conclusions on decisions affecting national healthcare. However, there is bias in the Polysomnography section where some generalizations are made that are arguable, not based on data, and clearly intended to degrade in-lab polysomnography and influence the reader.	It is notable that reviewer 4 states the introduction to the report is biased against PSG while reviewer 5 states that there is a bias toward PSG. We have made revisions to our discussion about both PSG and portable monitors to improve descriptions of both that we hope remove the suggestion that we have a bias toward or against any test.
4	Methods	The methodology as outlined seems appropriate, especially given the large search results. Although Figure 1 which is intended to show the general analytic framework is anything but clear. The criteria for setting review and study parameters are clear, reasonable, and based on input from a host of experts. Finally, the definitions for outcome measures seem appropriate and clear.	Analytic Frameworks are inherently “anything but clear” if they are at all complete. The figure has gone through many revisions. This is the best figure agreed upon with domain experts, methodologists, the Task Order Officer, others at AHRQ, and selected members of the Technical Expert Panel.

Reviewer[1]	Section[2]	Reviewer Comments	Author Response[3]
4	Methods	There are a couple issues that are somewhat bothersome. Most decisions concerning inclusion and exclusion criteria are described and generally justified within the context of difficult issues. However, some of the exclusion criteria are not easily understandable, such as excluding studies where >20% did not have OSA when evaluating treatment of OSA since 1 out of 5 participants could still be inappropriate. This may be a problem with clarity of description rather than the actual data. Another concern is the use of Bland-Altman plots to analyze sensitivity/specificity of OSA diagnosis based on the authors' notion that there is no gold standard for the diagnosis of OSA. Full polysomnography has been the de facto gold standard for diagnosing sleep apnea and is widely accepted as such, so it is not clear that Bland -Altman plots should be justified on this basis. These types of statements seem to reflect a bias of the authors which should not be detectable in an objective review paper. A final issue is that comments about the AHI and obstructive sleep apnea not being defined seem to confuse OSA with sleep apnea syndrome. The first is objectively defined and easily identified in respiration while the latter is a clinical judgment, but the authors do not make this distinction. Consequently their argument that OSA doesn't have a gold standard for diagnosis is not correct and confusing to the reader.	We have added to the Methods section about Population and condition of interest (page 16, top): This threshold (20 percent) was chosen arbitrarily to avoid excluding potentially relevant small studies that included some patients with conditions not of interest to the current report. This turned out to be a moot point since no eligible studies explicitly included patients with any of these conditions. Together with the TEP, we decided to follow the construct of our 2007 technology assessment on PSG. We came to agreement that PSG is an accurate measure of AHI and other (obstructive and nonobstructive) apnea measures but is not a definitive test for OSA (syndrome) since the definition of the syndrome includes clinical judgment and arbitrary thresholds. This condition of interest for this report is the clinical syndrome of OSA. We have expanded our description of these concepts in the introduction and added the above text to page 16 of the methods..

Reviewer[1]	Section[2]	Reviewer Comments	Author Response[3]
4	Results	The result section is quite extensive and appropriately detailed. The relevant characteristics of the studies are included and the figures are informative. Since additional analysis is minimal, the results are clear and easy to follow. The section seems well done, comprehensive and nicely presented.	Thank you
4	Discussion/ Conclusion	The numerous findings of this review are clearly summarized. Limitations of the included studies are discussed in a thoughtful and objective manner. Suggestions for future research are comprehensive and exceptionally well done. They are clear, specific, well thought out and in some cases well-detailed.	Thank you.

Source: www.effectivehealthcare.ahrq.gov
Published Online: October 25, 2011

Reviewer[1]	Section[2]	Reviewer Comments	Author Response[3]
4	Clarity and Usability	The report is well-structured and organized. The writing is generally clear and the results easy to follow. I believe the reviews provided and the conclusion that is drawn from the data can be used to help develop policy and practice decisions. For most of the newer OSA treatment modalities there is insufficient data to draw a conclusion about effectiveness. In contrast, screening questionnaires generally performed poorly. One area of results that should be used with caution is that of portable monitoring systems for diagnosing OSA. The systems included in the data analysis are quite variable in the parameters that are recorded, algorithms used, and the quality of the data obtained, so general conclusions about their usefulness for diagnosis of OSA may not be true for a specific system. The authors allude to this but there is not enough data available currently to evaluate portable monitoring systems by individually and none in head-to head-comparisons. The latter needed is appropriately included in the section for future research. Additionally, the review did not address data loss and the need for repeat studies in portable monitoring which is an important consideration for setting reasonable and cost effective policies. Overall, the review is comprehensive and the data is analyzed and summarized well. An important concern that was not addressed in the paper or included in the discussion is data loss and the need for repeat studies for patients undergoing portable monitoring at home. This is an important issue for insurance payers and those setting national policy. Perhaps that could be addressed in the discussion.	We agree that there is a paucity of data on direct or indirect comparisons between monitors. We address this evidence gap in the Discussion and future research needs section. The evidence on data loss has been included in the results (Pages 29&33-4 end of study description) and discussion (Page 139, bottom). A comment on direct and indirect comparison has been added to the discussion (Page 139, near bottom) and already exists in the Future research needs section (page 144).

Reviewer[1]	Section[2]	Reviewer Comments	Author Response[3]
4	Page 4 line 50	"AHI, the diagnostic value of which can vary from 5-to-20 events per hour." It is not clear what 5-20 events per hour are referencing as the AHI can vary from 0 to over 100. Clarification is needed.	Amended to: The minimum threshold to diagnose sleep apnea can vary from 5 to 20 events per hour.
4	Page 4 line 52	"signals that are interpreted by trained technologists and sleep physicians". This is incorrect as trained technologists do not interpret signals or sleep studies. This is only done by the physicians. The term trained technologists should be deleted or "interpreted" should be changed to "read".	Amended to: read by trained technologists and interpreted by sleep physicians
4	Page 5 line 6	states that PSG is not a definite test for OSA "due to a lack of standardized criteria for test parameters and the thresholds used to make diagnoses." This is not true as the AASM has set and published recording parameters and standards for PSG that are used throughout the world. The AASM has also published the ICSD2 that includes criteria for the diagnosis of obstructive sleep apnea.	We have added in the AASM criteria. But it remains true that among researchers variable criteria are used.
4	Page 5 line 3-7	"diagnosing OSA by PSG in a sleep lab is not, however, without considerable constraints including cost and interlaboratory variation in hardware and assessment methods. Therefore, it is acknowledged that PSG is not a definitive test to either diagnose or rule out OSA." In lab PSGs recordings are very similar if not indistinguishable regardless of the hardware used so it is not clear what the basis is for this statement. Moreover, the conclusion that "it is acknowledge that PSG is not a definitive test for OSA" cannot be drawn from this statement and seems to reflect the authors' strong negative bias.	We have toned this down by deleting: “Due to a lack of standardized criteria for the measures of test parameters and the thresholds used to make the diagnosis, it is difficult to diagnose or rule out OSA solely based on PSG.” And “it is acknowledged that PSG is not a definitive test to either diagnose or rule out OSA.”

Source: www.effectivehealthcare.ahrq.gov
Published Online: October 25, 2011

Reviewer[1]	Section[2]	Reviewer Comments	Author Response[3]
4	Page 5 lines 12-13	"the AHI, which is used as the single metric to define OSA for insurance companies and in clinical settings, can vary from night to night and does not take into account symptoms, co morbidities, or response to treatment." While this is true, it is not specific to PSG, but rather is a criticism of using the AHI for diagnosis of sleep apnea syndrome. Consequently it does not belong in this section.	This is a good point. We have moved this and similar text about AHI and severity of OSA out of the PSG section up to the prior introduction to Diagnosis.
4	Page 5 line 36	"If screening tests are, in fact, sufficiently predictive of the results of full sleep testing..." This is misleading to the reader as the terms "in fact" seems to indicate that this is true based on actual data but as the authors indicate there is no data and predictability is still unknown. A more accurate statement is "If screening tests are found to be sufficiently predictive..."	Amended as recommended.
4	Page 14 line 29	The exclusion criteria for Treatment of OSA are somewhat confusing. It states that "we excluded studies with >20 percent of study subjects without OSA..." More clarification needs to be given why data concerning treatment of OSA would include up to 20% of patients who did not have OSA.	See response above.

Reviewer[1]	Section[2]	Reviewer Comments	Author Response[3]
4	Page 134 line 31	The review combines data from various Type III and Type IV portable systems due in part to the paucity of data. As the authors indicate some will perform better than others so generalizations about the usefulness of portable monitoring to diagnose OSA many not be true for individual systems. However, the discussion states that "most of the tested portable monitors fairly accurately predict OSA, with high positive likelihood ratios and low negative likelihood ratios..." However including the qualification that the systems are quite different and some perform better than others should be clearly addressed in the discussion given the importance of this review for policy makers and given that moderate support is being given for the use of portable monitoring to diagnose OSA.	We have added some qualified language to the Discussion to address this point. Page 139
4	Page 52 line 3. & page 59 line 51	Wakefulness tests- The MSLT is not a test of wakefulness so this sections using this title are somewhat misleading and should be changed. A more appropriate title may be something like "Objective Daytime Tests "or "Objective Sleepiness/Wakefulness Tests"	We have changed all the “Wakefulness tests” titles to “Objective sleepiness and wakefulness tests”
4	Page 52 line 7	"The range of differences between the two interventions was large (-1.0, 2.4)..." However, it is not clear what these numbers represent or why this is considered a large difference.	We have changed the text to include units on this measurement (minutes) and modified the language to accurately reflect the difference (which we agree is not overly large). We have also added a meta-analysis.

Reviewer[1]	Section[2]	Reviewer Comments	Author Response[3]
4	Page 26 line 50	"The mean bias is the average difference between the AHI (or RDI or OHI) estimated with the portable device and the AHI measured by PSG." The AHI, RDI and OHI are generally correlated but they are not the same and actual values can be significantly different despite correlation. The data reported in this paragraph subsequently only include AHI and RDI comparisons. Where ODI values not ultimately used in any of the studies included? Also the impact on the comparisons of using different measures should be mentioned.	All of the studies on Type III monitors included RDI values (referred to as AHI in most manuscripts). When ODI was reported alongside RDI, both were extracted but only the RDI was used for analysis.
4	Page 26 line 40-54	It seems like the difference between PSG data and Type III devices is substantial with discrepancies ranging from 17.7-36 events/hr when the PSG AHI used for comparison ranged from 23-39.9. It seems inappropriate to consider either difference as low as stated "discrepancies can be as low as 17.7 events/hr and as large as 36 events/hr. It is more objective to say that the discrepancies ranged from 17.7 to 39.9 events/hr.	We have simplified the language and removed some of the "jargony" language and concepts.
5	Abstract	The abstract states the following: "Low strength of evidence that the Berlin Questionnaire diagnoses OSA . . ." I am not sure it is accurate to make a statement about a self-report questionnaire's ability to "diagnose" OSA. Perhaps the Berlin and similar questionnaires can be evaluated in their ability to be associated with the diagnosis of OSA. I would recommend rewording this phrase throughout the report.	You are correct. We have changed diagnose to predict or screen for, as appropriate, throughout the document.

Reviewer[1]	Section[2]	Reviewer Comments	Author Response[3]
5	Abstract	Issues related to OSA diagnosis and treatment are very different. To guide the reader, perhaps the abstract should reflect this difference. For example, 222 studies met eligibility criteria for the report, but what was the breakdown by diagnosis and treatment? I might recommend identifying the KQ in the abstract by diagnosis and treatment as well.	We have included the numbers of test, predictor, and treatment studies. Since the Key Questions are not included in the abstract, we instead have removed their numbers (eg, KQ1) in the Results section of the abstract. The summary sufficiently describes testing vs treatment.
5	Abstract	Conclusion KQ1 is that the portable monitors maintain a variable bias in estimating AHI. Is there anything more conclusive that can be said as a result of this very comprehensive report? Some ideas on improved data analytic procedures are offered below to offer new conclusions that are important for clinicians and researchers alike.	There is extremely limited space in the abstract. No changes were made here. Changes were made elsewhere, though, as suggested.
5	Introduction	Polysomnography. The sentence ". . .it is acknowledged that PSG is not a definitive test to either diagnose or rule out OSA" on page 5 is troubling in the context ofthe rest of the report. How can the test be called the "gold-standard" in the field, and in this report it is referred to as the "diagnostic standard" in use in clinical practice. It is further stated in this report that "the diagnosis of sleep apnea requires a comprehensive, technologist attended sleep study. . . " This paragraph requires a better discussion of PSG as a test for OSA, given its strengths and limitations, particularly to put that intial statement in better context for the reader. It really is an extraordinary statement. All of that said, the points made regarding its limitations are very important to include and relevant.	This section has been rewritten to improve clarity and to reduce perceived bias.

Reviewer[1]	Section[2]	Reviewer Comments	Author Response[3]
5	Introduction	There is a clear bias toward full, attended polysomnography (despite previous comment). The paragraph on portable monitoring is short and insufficient. The reality is that portable monitoring is being used and is approved for reimbursement. The report will only be useful to the extent that it adequately addresses concerns that our practitioners in the field are faced with each and every day. At the very least, rather than discussing the variability in parameters and data synthesis, discussion of the commonly accepted Types of diagnostic studies should be presented and discussed.	Please see our response to Reviewer 4 on the bottom of page 9 of this document. We have added in a new table (Table 1), originally from our 2007 report) that summarizes the different monitor types.

Reviewer[1]	Section[2]	Reviewer Comments	Author Response[3]
5	Introduction	Further consideration of the reporting of Type III and Type IV devices. (1) Type III and Type IV devices should not be lumped together in making conclusions. (2) if the authors are agreeable that Type III, but not Type IV, provide the physiological data required by the AASM 2007 criteria for scoring apneas and hypopneas, it is this reviewer's belief that a separate analysis should be provided on those Type III devices that are able to provide that data. This would be the most relevant comparator to full PSG, and quite frankly, the most important analysis of portable monitoring. The authors even in the full PSG introductory paragraph discuss the limitations of full PSG in terms of unstandardized criteria. However, our field must meet the needs of our patients each and every day, so to ignore current working criteria (ie, AASM 2007) does a disservice to our community and this report is not as relevant as it could and needs to be. To that end, all reporting related to KQ1 and portable monitoring should be first separated by Type III and Type IV devices, and then the Type III devices should be further subdivided by those devices that do and do not provide physiological data that allows for scoring according to commonly accepted criteria (e.g., AASM 2007). The main point for doing this is to have the report provide more focused analysis and results that are directly relevant to clinicians and researchers. I would imagine the analysis may have to identify the specific sensors used in each study or by each device. It may be that the more focused analysis would have to identify the specific Type III device utilized. From Table A, the reader should be ableto see the results of the most relevant Type III devices based on studies with the grade of A.	We reviewed and summarized in the Results for KQ 1 how many studies evaluated each of the portable monitors. Very few monitors have been evaluated by even 2 or 3 studies. No study directly compared 2 or more monitors. We did not think it would be fruitful to subcategorize monitors beyond monitor Type. We also did not believe the indirect comparisons across studies would be an appropriate way to compare monitors. Pages 28, 33, and 35.

Reviewer[1]	Section[2]	Reviewer Comments	Author Response[3]
5	Introduction / Results (KQ1)	In an ideal situation, a Type III device would measure one airflow, one effort and one oximetry channel. If full PSG uses the identical channels, but has the added benefit of sleep measurement, one would think the numerator would be similar, but that the variability would come from the denominator of the AHI. Of course, when measured on different days, some of the variabilty would come from natural night-to-night variability. The report would benefit from acknowledgement of this night-to-night variability and from reporting on any PSG results that directly report on this variabilty. If it were possible to "adjust" the portable/PSG variablity rates based on this additional source of variabilty, all the better. Further, some discussion about study design would be relevant as well: clearly, nights where both devices were worn simultaneously would provide the best comparison (and control for nightly variability; but would obviously come at a cost).	A discussion of night to night variability has been added to the Discussion section (Page 145). The studies did not allow us to adequately assess any issues related to night to night variation. However, it is important to note that simultaneous testing is not a real-world test of portable monitoring.
5	Results (KQ1)	It may not be realistic at this point to substantially alter the report. If not, my recommendation would be to consider the above as an additional subanalysis under KQ1.	The 2007 Technology Assessment of Home Diagnosis of Obstructive Sleep Apnea-Hypopnea Syndrome conducted by the Tufts Evidence-based Practice Center found that portable monitors in the home setting performed poorly as compared to portable monitors in the lab setting, where the measurement would be simultaneous. We did not reevaluate studies included in the 2007 report. Since the 2007 report, only one study among Type III monitors and 3 studies among Type IV monitors were performed in the home or community setting. However, 9 studies (3 among Type III and 6 among Type 4) were done using the same device in both lab and home settings. Among these 9 studies, there was no difference between the two settings in the range of mean bias, sensitivity and specificity reported. This has been added to the report results (Pages 29, 30, 33-35) and the discussion (Page 139, bottom) .

Source: www.effectivehealthcare.ahrq.gov
Published Online: October 25, 2011

Reviewer[1]	Section[2]	Reviewer Comments	Author Response[3]
5	Introduction	Pretesting Questionnaires. Here, questionnaires are appropriately referred to as prescreens for further testing or treatment. The abstract should be modified accordingly.	We changed the abstract to: Berlin Questionnaire is able to prescreen patients with OSA
5	Introduction	Is the Stanford Sleepiness Scale still used? There is much anecdotal information about its lack of validity as a self-report tool (ie, patient may rate self alert while eyes are shutting). If it were used as a clinical observation tool or rating scale, it may have merit. As a self-report tool, I would drop its mention from the report.	We agree and have removed the SSS from the Introduction

Reviewer[1]	Section[2]	Reviewer Comments	Author Response[3]
5	Introduction (Treatment)	CPAP. I would recommend that the report not offer a definition of compliance that has nosupport in the literature. Even Dr. Weaver, who co-authored the original paper of no more than 40 subjects which happened to show a bimodal distribution, indicated it was not appropriate to derive a compliance definition from that dataset. There is now increasing evidence that more CPAP use is associated with improved outcomes. Whether there is some minimal threshold, maximal threshold or curvilinear relationship is not currently known. What the evidence suggests so far is that more use is better, and all indications thus far are for a linear relationship. Why not simply state that there is no formal definition of compliance, but that the evidence to date suggests that more use is associated with better outcomes? It would be wrong for this report to imply that 3.75 hours of CPAP use on 70% of the nights is non-compliant, when in just may be that that patient has decreased their risk of cardiovascular problems, reduced their sleepiness level and increased their quality of life. Again, the comprehensiveness of this report allows it to approach common conceptions in a novel way and offer new thoughts and directions when possible.	We have removed the example definition of compliance and replaced with a general definition of use only part of the night or some nights. In the Results section, when summarizing studies, we used each study's definition of compliance.

Reviewer[1]	Section[2]	Reviewer Comments	Author Response[3]
5	Introduction (Treatment)	2nd and 3rd paragraphs. It is implied that value of psychoeducational interventions remains "unclear" but no such statement is made in paragraph 3. In fact, paragraph 3 is written with a bias toward technological improvements improving adherence. Given that the aim of the report is to review the evidence, any interpretive comments should be withheld in these paragraphs. To that end, paragraph 2's last sentence should be removed, and another sentence describing the full range of psychoeducational intervention efforts should be made to adequately describe the full range and format of such interventions. The sentence in paragraph 3 " other cpap machines. . . designed to improve comfort and therefore compliance" should be altered or removed. In this reviewer's opinion, the evidence that shows that any technological improvement results in increased adherence is poor at best. Perhaps such statements should be made in a hypothetical context, as a way to lead in to the results of this report.	We have removed the term “advancements” and replaced with more neutral words (features, modifications). We have also added an equivalent section to this paragraph: “The value of the modifications to CPAP, however, remain unclear.”
5	Introduction (Treatment)	Paragraph 4 assumes that there are known correlates of compliance that can be measured prior to treatment. It is incumbent on the authors provide the evidence that support this paragraph. It is this reviewer's belief this is very difficult to do, whether in the OSA field or outside of it in other disease conditons. There is also an assumption that any interventional efforts would require extra resources of both patients and providers, and perhaps of the healthcare system. This may not be the case. It just may be the case that improved efficiencies are what is needed. I would recommend being cautious with the assumptions in this paragraph.	We’ve rephrased, principally by saying that improving compliance *may* require health care resources etc. We think the paragraph is appropriate and reasonable without the need to make any unrealistic assumptions.

Reviewer[1]	Section[2]	Reviewer Comments	Author Response[3]
5	Introduction (Treatment)	An additional sentence on the reason for CPAP titrations being suboptimal is warranted. The last paragraph in this section is important to include, as some of this is being done in the field. Further, there is increasing evidence that pressure requirements can vary significantly once someone is on CPAP. So, even with optimal in-lab pressure settings, once the patient is in the real world, no matter how valid or reliable an initial pressure setting is, appropriate follow-up and monitoring is critical.	We agree. We have added “Close followup and monitoring of CPAP pressures is important in all patients regardless of how the CPAP level was determined.”.
5	Introduction (Treatment)	The dental and mandibular device paragraph is short and to the point. It does a very good job of covering this important area.	Thank you.
5	Introduction (Treatment)	The surgery paragraph would benefit from a sentence or two on the outcomes expected with surgical techniques.	We agree and have added “In general, the goal of surgery is to remove the anatomic obstruction and to relieve symptoms.”
5	Introduction (Treatment)	The preliminary nature of the work on pharmacologic agents would be important to include.	We don’t agree that this is necessary to make explicit. This is implied by their being lumped into Miscellaneous treatments and, more importantly, this is a conclusion to make after reviewing the Results.
5	Methods	Literature search. The search is a top-down strategy that is reliant on the system that articles are appropriately labeled and categorized. The potential limitation of such an approach should be mentioned. Bottom-up strategies where one takes the broadest possible approach and relies on human review is more time intensive, but less likely to be reliant on the systemThe fact that the search was supplemented by reference list review is very good.	We don’t think it is necessary or even appropriate to claim that electronic searching is limited. A bottom-up strategy that attempted to overcome misappropriate labeling or categorization is not only time consuming but would be impossible. Every full text article would need to be screened. This is a much greater limitation (that does not need mention either.
5	Methods	Why does hemoglobin A1c play such a prominent role in this report? The fact that it is listed as one of 2 intermediate outcomes is surprising. One would expect to see blood pressure here, for example.	We do not believe Hb A1c has such a prominent role in the report. Blood pressure is also evaluated. Both were rarely evaluated by the eligible studies and their importance is downplayed. They do not appear in the overall summaries, which we believe is appropriate.

Source: www.effectivehealthcare.ahrq.gov
Published Online: October 25, 2011

Reviewer[1]	Section[2]	Reviewer Comments	Author Response[3]
5	Methods	Portable monitoring. It appears that portable monitoring may have been lumped together, per page 14. This would be a mistake. Perhaps the portable devices should have been organized by Type per comments above.	All interventions were sought simultaneously. The devices were analyzed and reported separately.
5	Methods	Treatment compliance studies. In all likelihood, one will find studies that are interventional in nature (ie, primary outcome of interest is adherence, based on some sort ofeducational or technological intervention); studies of correlates of adherence; or studies where the primary focus is not adherence at all, but it happened to be measured. How did the report distinguish these, and other types, of studies, if at all?	We did not have particular interest in the primary focus, purpose, or outcome of interest to the original study researchers. Instead we focused on the analyses and results reported. Thus, we didn't distinguish the studies. We evaluated them for whether they addressed our Key Questions (as an outcome for Key Questions 5 (usually not 1º outcome) and 7 (usually 1º outcome) and whether there are predictors of compliance (Key Question 6, usually, though not always, a 1º outcome)).
5	Methods	Why were set cut-offs of a certain sample size utilized, and what were their justification? On the one hand this is a good idea, but on the other hand well done studies may not have been included.	We have added the explanation (p 19, near bottom): Sample size thresholds were chosen based primarily on practical consideration of available resources and time balanced with the likely amount of available literature.
5	Methods	Quality assessment. The use of a simple, straightforward system is prudent. What role did industry funded studies play? This report would be remiss if it did not look at funding source.	We have added several sentences to the Limitations sections of the Executive Summary and Chapter 4 about the concern of industry support on possible publication bias.
5	Methods	Grading body of evidence. This also is a strength of the report.	The EPC program agrees.

Reviewer[1]	Section[2]	Reviewer Comments	Author Response[3]
5	Results (Weight loss)	Given the importance of weight loss for a relatively large number of OSA patients, it might be interesting to expand a bit on the finding that "all studies found statistically significant correlations between the amount of weight loss and the magnitude of AHI decrease." For example, what was the average correlation coefficient, and if it were possible to put this into a "average 5lb weight loss is associated with x point reduction in AHI" I think that would be helpful for the reader.	Statistics given on correlation between weight loss and AHI were not consistent or clear between studies. This sentence has been removed from the report for the sake of clarity and accuracy.
5	Results (KQ6)	Predictors of compliance. I was surprised to see that only 5 studies were included in this analysis. Table A KQ6 might benefit from the following label "Predictors of CPAP compliance" to distinguish it from compliance with other treatments.	These were the only studies that met (arguably stringent) criteria. In fact this section covers compliance with (theoretically) all treatments (in practice, CPAP and MAD). We have added the word treatment, but did not further separate CPAP from MAD.
5	Results (KQ6)	CPAP compliance intervention studies: Was self-reported compliance allowed? If so, this should be mentioned in the limitations section of the report.	We have added a paragraph to the Limitations section of Chapter 4.
5	Page 124	Page 124 reviews the Richards 2007 study and reports a positive finding. Table A on page xxxi does not reflect the positive finding. Unfortunately, this level of review was not possible given the sheer volume of the report. All that worked on the report should be commended for a tremendous report!!	Thank you. Table A was corrected.
5	Discussion	The future research recommendations are well thought out and presented.	Thank you.

Source: www.effectivehealthcare.ahrq.gov
Published Online: October 25, 2011

Reviewer[1]	Section[2]	Reviewer Comments	Author Response[3]
6	General	This is a well-written and thorough systematic review of the comparative effectiveness of diagnostic strategies and treatments for obstructive sleep apnea in adults. The authors have clearly defined the target population for studies included for analysis with minor exceptions (details below). The key clinical questions are well stated, although there were some additional questions that might merit consideration (see below). Inclusion and exclusion criteria implemented appeared justifiable and the search strategy was explicit. Statistical methods employed appeared to be appropriate. The key messages appear to be explicit and applicable with a nice presentation of data in tables and figures. On rare occasion, there may have been some studies that were not considered (detailed below). This is a clinically meaningful report that can be used in policy decisions and practice decisions and to support the conduct of future studies due to limited evidence in a number of areas.	Thank you.
6	General	1. What is the level of sleep apnea severity that generally merits treatment? • What if any evidence is there to support what level of sleep apnea merits therapy? Does mild OSA need to be treated at all? The authors might consider bringing this out at the very lease in their section on areas of future research.	We have made revisions to the 2nd bullet under Future Research/Treatment: • Primary studies on the modifying effects of different patient characteristics, baseline disease severity, and other relevant parameters on various treatment outcomes should be undertaken so that treatment options can be optimized for or can be focused on patients with specific profiles, thus maximizing treatment benefits. Studies should be large enough, of sufficient duration, and bear minimal loss-to-followup rates to allow meaningful subgroup or regression analyses.

Reviewer[1]	Section[2]	Reviewer Comments	Author Response[3]
6	General	2. What AHI level should represent clinically significant disease? • This is a variation on the above question. Current studies utilize a consensus based definition to define mild, moderate, severe disease. Such definition would ideally be linked to an outcome whether an intermediate or clinical endpoint. The authors might consider bringing this out at the very lease in their section on areas of future research.	We chose to defer this issue to guideline development workgroups and not do discuss it in our report
6	General	3. What is the evidence base for the appropriate definition of a hypopnea? • This is a major area that needs to be considered, given the variety of definitions implemented in studies. CMS has adopted the position that a hypopnea is a respiratory event of certain magnitude reduction in association with a 4% drop in saturation. Similar respiratory events in association with arousals, but with a lesser drop in oxygen saturation are not considered to be respiratory events under CMS criteria. Nevertheless, the medical community, used these events with arousals to treat patients. This is a key clinical question that should be considered, particularly for future areas of research.	This was not specifically a topic covered by the Key Questions. We did not review this topic and have made no changes to the report.

Reviewer[1]	Section[2]	Reviewer Comments	Author Response[3]
6	Results	1. Portable studies vs. PSG – The authors conclude that there is moderate strength of evidence that type III and IV monitors may have the ability to predict an elevated AHI. While I generally agree with their conclusions, I believe that this conclusion should be tempered by often poor description of the samples studied in many of the publications. In particular, the utility and validity of portable monitoring in specialized populations such as those with chronic lung diseases, congestive heart failure, on pain medications, or neurologic disorders, where concerns for central sleep apnea may be more of a concern and the use of PM may be inappropriate should be clearly stated.	The applicability issue added to Tables A & 4 and to Discussion (page 137)
6	Results	1. Was the study by Richards et al (Sleep 2007;30:635) considered for the section on CPAP interventions to improve compliance?	Yes. Reference 302.
6	Introduction	1. On page 35, line 9, the authors references study by the “American College of Chest Physicians". The study referenced was performed by the National Sleep Foundation and published in the journal, Chest.	Thank you. This has been corrected / clarified on Page 1
6	Introduction	2. On page 35, the first paragraph last sentence - hypopneas and apneas also result in intermittent arousals from sleep. They should be noted in addition to the oxygen desaturation.	This is in the next paragraph of the Introduction. But we have added to the first Background paragraph in the Executive Summary.
6	Introduction	On page 35 line 17, the authors indicate that airway obstruction leads to a disruption of rapid eye movement sleep. It would be more precise to indicate that airway obstruction leads to a disruption of sleep.	“Frequent arousals from sleep” = disruption of sleep. Our local expert thought it was important to emphasize REM.

Reviewer[1]	Section[2]	Reviewer Comments	Author Response[3]
6	Introduction	On page 3, line 43 - the authors indicate that there is disagreement about the type and level of respiratory abnormalities that should be used to define sleep apnea, particularly for patients who have hypopnea episodes rather than apnea. I am not sure that there is disagreement in the literature about the issue hypopnea episodes rather than apneas. Can the authors be more precise in their intent? It would make sense to provide a reference here.	This section has been rewritten based on others' comments.
6	Introduction	It should be made clear that the authors are focusing on obstructive sleep apnea only. This is done to some extent on page 4 of the introduction, but should come out more clearly in the executive summary. Furthermore, it may be useful for the authors to indicate that there review does not encompass the issue of "complex sleep apnea". I find mixed sleep apnea to be a poor term to use as it is generally reserved for describing a respiratory event with both central and obstructive features. On page 4, line 12, I would insert, "events with" between involves and features.	It is stated quite clearly and explicitly in several locations in the Background and Methods that this report focuses on OSA. Mixed sleep apnea, in our reading, seems to be the more common term. "events with" added. Page 2.
6	Introduction	On page 4, line 23 between the words, "but manifest" should be "but also manifests".	We have revised.
6	Introduction	On page 4, line 30 - it is unclear what the word "It's" is referring to. Please be more clear.	We have revised to "The report's"
6	Introduction	On page 4, line 37 -it should be noted that upper airway obstruction leads to both intermittent hypoxemia in nighttime arousals. In addition the phrase, "once they are unable to breathe" is unclear and perhaps unnecessary.	Both events are already included. Phrase deleted.

Reviewer[1]	Section[2]	Reviewer Comments	Author Response[3]
6	Introduction	In the preamble to the diagnosis, the authors briefly describe that the pathophysiology of sleep apnea is due to anatomic abnormalities and decreased pharyngeal muscle tone during sleep. It would be perhaps more accurate to indicate that anatomic abnormalities, decreases in pharyngeal muscle tone and insufficient neuromuscular responses to airway obstruction likely contribute to the pathophysiology of the disorder.	Added on Page 2
6	Introduction	Under this section polysomnography on page 4, line 50 - the authors state, the diagnostic value of which can very from 5-20 events per hour". It is unclear what is being referred to here. Are the authors attempting to define what the lower limit of disease definition has been? Please clarify.	This has already been clarified based on other reviewers' comments.
6	Introduction	On page 4, line 55 - the authors should indicate that both chest and abdominal wall monitors are used to evaluate respiratory movements.	Added on Page 3
6	Introduction	On page 5, the authors discuss the use of the AHI is used as a single metric to find OSA for insurance companies. It may be worthwhile for them to point out the controversy with respect to the use of AHI which does not take into the account hypopnea events associated with arousals versus the use of the RDI, which would take into account respiratory events associated with arousals.	This may be an important topic for a guideline workgroup, but we don't think it's sufficiently relevant for this review.
6	Introduction	On page 5, under pretesting questionnaires and other tests, it may be useful for the authors to indicate that the Epworth sleepiness scale in Stanford sleepiness scale although used commonly are not necessarily specific to the diagnosis of obstructive sleep apnea.	Added on Pag3 4

Reviewer[1]	Section[2]	Reviewer Comments	Author Response[3]
6	Introduction	On page 6, line 52 under the section on continuous positive airway pressure the authors discuss technologies such as C-flex™ . It may be appropriate to be more generic about the use of this term as there other companies that have similar technology (e.g. EPR by Resmed, Inc).	See addition on Page 6. There doesn't seem to be a good generic name for these devices.
6	Introduction	On page 7 the authors discuss when a split night is indicated - please provide a reference for this statement.	This sentence was deleted.
6	Introduction	It may be more appropriate to consider the use of the term limited channel testing in place of portable monitoring.	No change was made
6	Introduction	The authors indicated that they considered miscellaneous interventions for the treatment of obstructive sleep apnea. One area that was not mentioned with some publications is the area hypoglossal stimulation.	We did not find any such studies that met criteria. Most likely there are no RCTs.
6	Results	Portable Devices vs. PSG's References to Figures 1.1 and 1.2 in text (p27, line 38) refer to areas of shading on the figures that are not apparent.	This has been corrected
6	Results	Results – Questionnaires vs. PSG's Page 65, line 16 – the authors indicate that 53 to 211 participants were studied. According to table 1.4.1 it would suggest that there were 53 – 2127 participants in any given study. I suspect the authors meant that only 53-211 of the 53-2127 participants evaluated had sleep study testing? Please clarify.	Done

Reviewer[1]	Section[2]	Reviewer Comments	Author Response[3]
6	Results	Results – Clinical prediction rules vs. PSG's Page 69, lines 14-26 – The authors summarize the findings of Onen 2008 in this section. The paragraph would be more readable if the sensitivity and specificity findings are reported in order for the >= 2 snoring episodes, then >=3 snoring episodes, and >=5 snoring episodes, rather than as currently presented.	Thank you. We agree and have fixed.
6	Results	Results – Preoperative screeing The authors indicate only 2 studies met selection criteria. Based on their criteria, it seems that the study by Gali et al, Anesthesiology 2009;110:869, would be included. Was the study considered?	Gali did not compare screening with another strategy. The preamble to KQ 3 has been clarified. Page 47.
6	Results	Results – Key Clinical Question 5 – comparative effects of different treatments for OSA The authors have a category called Wakefulness Tests(p84 and 92), that reports studies that use an MSLT or MWT as an outcome. I would consider renaming this section as an MSLT is not a "Wakefulness Test".	This has already been corrected based on other reviewers' comments.
6	Results	Comparison of CPAP vs. sham CPAP(p91) On line 22, the authors report the sample size range as 25 to 101. It would be informative and parallel to previous sections if they also include the total sample size of studies included for analysis.	We have added this in. This was the only relevant place this was missing.

Reviewer[1]	Section[2]	Reviewer Comments	Author Response[3]
6	Results	Results - Comparison of oral and nasal CPAP (p99) The authors report on comparisons of oral and nasal CPAP. It is important to note, however, that one of the studies included in this section by Mortimore in 1998 was a comparison of nasal mask to full face mask (not an oral mask). This study should not be included as part of this analysis and may deserve a separate paragraph under a more encompassing heading or separate heading.	We agree. We have not restructured, but clarified that Mortimore used a full face mask. This did not change conclusions. Changes made on pages ES-8 and 70.
6	Results	Results – Comparison of Bilevel CPAP and CPAP (p108) The authors heading is inaccurate. Bilevel positive airway pressure is not typically referred to as "Bilevel CPAP". The term ". Bilevel positive airway pressure" or "bi-level PAP" would be preferable. The term is also used on p 109, line 47.	We went back and forth on the terminology and decided for the sake of consistency and clarity to stick with CPAP in preference to PAP. The term Bilevel CPAP is also used by others.
6	Results	One of the papers used for analysis in this section (Piper 2008) is a study of CPAP vs. bilevel support in patients with obesity hypoventilation syndrome rather than just obstructive sleep apnea. This is a different patient population than that considered in this systematic review and should be considered for exclusion from this analysis.	We did not exclude obesity hypoventilation syndrome. After review of the methods by our local expert, we concluded that these patients have obesity-related OSA. We have made it more explicit that the patients have obesity hypoventilation syndrome. Page 79.
6	Results	Results – comparison of Cflex and CPAP (p110) Were there any appropriate studies using expiratory pressure relief (EPR) that is utilized by Resmed, Inc.? These would be useful studies to include if appropriate and available.	No. We have clarified this in the Introduction. Page 6.

Reviewer[1]	Section[2]	Reviewer Comments	Author Response[3]
6	Results	Results – TRD vs. TRD (p123) It is unclear what is compared in the heading. The authors should make explicit in the heading that the devices compared are a TRD with and without suction.	Thank you. The header has been revised (with versus without suction) and the suction device has been described more fully. Page 94.
6	Results	Results – Comparison of weight loss interventions and control interventions (p134) The authors report on 3 parallel trials of weight loss interventions. Would the study by Smith et al, Ann Intern Med 1985; 103:850 have met study inclusion criteria, the study indicates that patients were randomly assigned to weight loss or no weight loss. If not why not?	Smith et al. 1985 was rejected at the abstract level. This is why it does not show up in the list of excluded studies. N<10 in the control arm (8), thus study was excluded.

[1] Peer reviewers are not listed in alphabetical order.
[2] If listed, page number, line number, or section refers to the draft report.
[3] If listed, page number, line number, or section refers to the final report.

Source: www.effectivehealthcare.ahrq.gov
Published Online: October 25, 2011

Reviewer Name[1]	Reviewer Affiliation[2]	Section[3]	Reviewer Comments	Author Response4
Burton Abrams		Executive Summary	The second sentence grossly underestimates the prevalence of OSA in the US. With reference to [1], it claims that 4% of men and 2% of women have OSA, leaving out the fact that those percentages are associated with daytime hypersomnolence, without which the percentages given in [1] are 24% of men and 9% of women. _____ [1] Hiestand DM, et al. Prevalence of Symptoms and Risk of Apnea in the US Population. Chest 2006,130:780-6.	More complete and better data are now reported in the introduction.
Anne Abreu	Ventus Medical		Ventus Medical respectfully requests that the AHRQ consider one additional study in the review. The study was included in abstract form in Ventus Medical's Scientific Information Packet submission to AHRQ on August 3, 2010. The manuscript is now peer-reviewed and was accepted by SLEEP on November 13, 2010. It is currently e-published at the following link: http://www.journalsleep.org/AcceptedPapers/SP-329-10.pdf This study is of particular relevance to the AHRQ review as it is a Randomized Controlled Trial of PROVENT Therapy vs Sham Device. An updated summary table of the findings is attached. Ventus Medical believes that PROVENT Therapy and this study are important to include in the AHRQ review, and based on the Study Selection and Eligibility section of the draft, believes that the study will meet all eligibility criteria for inclusion.	Unfortunately, we are not able to include this study in the current report as it has not been/was not published in time. It should be reviewed in any future update of this report.

Reviewer Name[1]	Reviewer Affiliation[2]	Section[3]	Reviewer Comments	Author Response4
American Sleep Apnea Association	American Sleep Apnea Association		It is unfortunate that, at the present time, so many of the answers are not supported by clear data, and that the AHRQ report finds most of the key questions to have low or insufficient evidence to answer. These results should be used to encourage obtaining data, and not be interpreted as there being no answer or that the question is not important. Those preparing the review indicated a significant lack of clinical outcome data to support anything but “all cause mortality” in the patients with the severest forms of the condition. Issues such as improvement in quality of life or neuro-cognitive functioning, both very important outcomes, are supported by little evidence. Both with respect to Positive Air Pressure (PAP) and Mandibular Advancement Devices (MAD) there was an insufficiency of evidence to address, which patients might benefit most from treatment. The insufficiency of the evidence extends to evaluating the comparative effectiveness, the purpose of the study of the three different treatment modalities – PAP, MAD and surgery. Despite these negative results there is confirmation of certain aspects of OSA diagnosis and treatment – · Type III and IV “limited” channel studies are generally accurate to diagnose OSA; · An Apnea Hypopnea Index (AHI) greater than 30 events/hr is an independent predictor of all cause mortality; · Given the large magnitude of effect on the important immediate outcomes, such as AHI there is moderate evidence to show that PAP is an effective treatment for OSA, · The same is true of MAD with respect to OSA in patients without co morbidities or excessive sleepiness. The ASAA agrees with the conclusions of the CER that additional research needs to be done if the benefits of treatment using the various	No changes made based on these comments.

Reviewer Name[1]	Reviewer Affiliation[2]	Section[3]	Reviewer Comments	Author Response4
Anonymous Reviewer 1		What is the effect of pre-operative screening for sleep apnea on surgical outcomes?	The best study looking at OSA as a risk factor for perioperative complications is for knee and hip replacement at Mayo Clinic. The citations is Gupta et al. Mayo Clinic Proceedings 2001. There are multiple studies on a wide variety of screening methods but none that compare the incidence of complications under each.	We did not include (accept) retrospective (case control) studies. If this is the best study, then we are correct that the evidence is inadequate.

Reviewer Name[1]	Reviewer Affiliation[2]	Section[3]	Reviewer Comments	Author Response4
Anonymous Reviewer 1		What is the comparative effect of different treatments for obstructive sleep apnea (OSA) in adults? a. Does the comparative effect of treatments vary based on presenting patient characteristics, severity of OSA, or other pre-treatment factors? Are any of these characteristics or factors predictive of treatment success? Characteristics: Age, sex, race, weight, bed partner, airway and other physical characteristics, specific comorbidities OSA severity or characteristics: Baseline questionnaire	The multiplicity of definitions of hypopnea and changes over time make such comparisons impossible.	It is a complicating factor, but our expert and the TEP did not think that changing definitions of OSA or of AHI were sufficiently important to preclude analysis or meta-analysis.

Reviewer Name[1]	Reviewer Affiliation[2]	Section[3]	Reviewer Comments	Author Response4
Anonymous Reviewer 1		What are the pre-treatment patient-level characteristics associated with failure to adequately treat OSA?	It depends what you mean by adequately treat. Please see comments from Ed Weaver (Weaver EM. Sleep apnea devices and sleep apnea surgery should be compared on effectiveness, not efficacy. Chest 2003) describing the difference between efficacy and effectiveness, a critical factor that is often ignored by many in the treatment of obstructive sleep apnea. For example, CPAP has high efficacy but variable effectiveness limited by compliance. Surgery has efficacy that is lower in most cases but effectiveness equal to efficacy because there are no issues of compliance.	We do not believe that we have conflated or confused efficacy and effectiveness.
Anonymous Reviewer 1		What is the effect of interventions to improve compliance with device (CPAP, oral appliances, positional therapy) use on clinical and intermediate outcomes?	There are multiple studies of nasal surgery (actually more than medical treatment of nasal obstruction) suggesting that this can improve compliance with CPAP. FYI, your reference (in the background) for perioperative complications after OSA surgery is not a good one. I am not sure how and why you would use this review article, which does not even review the literature well on this subject. It would be much better to have primary research than review articles. Instead, you might want to use citations of large cohort studies, including the largest one that included two papers on the incidence of complications and risk factors: Kezirian EJ, Weaver EM, Yueh B, Deyo RA, Khuri SF, Daley J, Henderson W. Incidence of serious complications after uvulopalatopharyngoplasty. Laryngoscope, 03/2004;114(3):450-3.Kezirian EJ, Weaver EM, Yueh B, Khuri SF, Daley J, Henderson WG. Risk factors for serious complication after uvulopalatopharyngoplasty. Arch Otolaryngol Head Neck Surg, Oct/2006;132(10):1091-8.	Any surgical studies (included those of nasal surgery) that met our broad criteria would have been included. We required the studies to be randomized trials. Chung 2008 is a systematic review, although not the best-reported systematic review. We believe that citing a systematic review is less biased, and more likely to be accurate, than citing individual studies. The two references offered here are specific to complications after UPPP. However, in this section (Preoperative testing) the question is whether it is of value to find undiagnosed OSA in patients undergoing surgery (anesthesia), not complications related to surgery for OSA.

Reviewer Name[1]	Reviewer Affiliation[2]	Section[3]	Reviewer Comments	Author Response4
Kim Brew	Medtronic	Structured Abstract	"Background" states that the diagnosis and treatment of OSA are cumbersome, resource intensive and poorly tolerated, yet the report describes the effectiveness of portable monitors and polysomnography (PSG) in diagnosing OSA, and of continuous positive airway pressure (CPAP) in treating it. The review also describes positive responses to various surgeries for OSA when CPAP is ineffective. So from its outset, the review should adhere to its purpose of providing a clear perspective on the management of OSA. • "Purpose" and "Data Synthesis" refer to screening and routine preoperative testing for OSA. Since these forms of testing are not covered by healthcare insurers, including Medicare, we believe that it should not be addressed in this report. (See also Key Questions [KQs] 2, 3 and 4 below.) • "Data Synthesis" states that trial evidence is insufficient regarding the relative value of most other OSA interventions, including surgery. We believe that this conclusion is too broad-sweeping, for prospective surgical studies based on anatomic site(s) of OSA have been demonstrating the efficacy of various operative procedures, such as uvulopalatopharyngoplasty (UPPP) and tongue-base suspension.	This CER was not designed or written specifically for Medicare or healthcare insurers. Nor were the lists of interventions or outcomes determined based on current coverages or preferences of payors. We included all interventions and outcomes of interest to OSA experts (clinicians) and patients. We made a point of saying "trial evidence" since this was the type of evidence included in this CER. We agree that there are many noncomparative surgical studies, but we included only a few of these for limited evaluation.

Reviewer Name[1]	Reviewer Affiliation[2]	Section[3]	Reviewer Comments	Author Response4
Kim Brew	Medtronic	Executive Summary	The “Executive Summary” introduces the reader to the purpose of a “Comparative Effectiveness Review,” specifically of different medical interventions. We believe that the “Background” to this summary should make clear that treatment of OSA is initially medical, usually CPAP, and if it is ineffective, various surgical options can be considered. Additionally, this would indicate to the reader of this CER why prospective, randomized studies comparing the effectiveness of medical with surgical interventions for OSA have not been performed. Also in the “Executive Summary,” we disagree with statement that confounding the diagnosis and treatment of OSA “is the great degree of clinical uncertainty that exists regarding the condition, due in large part to inconsistencies in its definition.” We believe that this statement and the foregoing definition of OSA should be clearer: “OSA is a syndrome characterized by repeated episodes of apnea and hypopnea due to repetitive narrowing or collapse of the pharyngeal airway during sleep, resulting in sleep fragmentation.”1 Further, “for determination of degree of sleep fragmentation, most sleep physicians assess the hypopnea-apnea index (AHI), or the average number of hypopneas plus apneas per hour of sleep over a period of several hours. Recommended AHI criteria for diagnosing OSA vary, but appear to favor either an AHI of 15 or more, or an AHI of 5 or greater and 14 or less with documented OSA symptoms.2,3,4” In light of these suggested revisions, we also disagree with the associated statements that “ongoing debate surrounds what type and level of respiratory abnormality should be used to define the disorder as well as the what is the most appropriate diagnostic method for the detection of OSA.” As we have indicated above, OSA has a clear, generic definition. The debate to which the authors are referring is not how to define OSA, but rather how to determine where narrowing or collapse is occurring, particularly	We are not providing a guideline or recommendations for management. We state that The most common first line therapy is use of continuous positive airway pressure (CPAP) devices during sleep. We have added to the description of Surgery that surgery is used for those with prior treatment failures with noninvasive techniques. Different definitions of OSA (or at least different criteria for OSA treatment) abound. Our clinical expert stands by our description.

Reviewer Name[1]	Reviewer Affiliation[2]	Section[3]	Reviewer Comments	Author Response4
Kim Brew	Medtronic	How do different available tests compare in their ability to diagnose sleep apnea in adults with symptoms suggestive of disordered sleep? a. How do these tests compare in different subgroups of patients, based on: race, gender, body mass index, existing noninsulin dependent diabetes mellitus, existing cardiovascular disease, existing hypertension, clinical symptoms, previous stroke, or airway characteristics?	The airway characteristics that are associated with OSA are not addressed. In parallel with surgical developments for OSA has been selection of diagnostic modalities to assess the complex anatomic and physiologic features of static pharyngeal airway narrowing and dynamic pharyngeal airway collapse, to determine what operative procedures are indicated.5,6,7,8 The most commonly used and effective modalities are fiberoptic nasopharyngeal endoscopy with the Mueller maneuver, nasopharyngeal manometry, cephalometry, and computed tomography (CT) of the pharynx.9,10,11,12,13 For Fujita,14 these modalities enable the classification of obstruction as retropalatal (type I), retrolingual (type II), or both (type III), and similarly for Riley,15 they enable the 3 classification of obstruction as oropharyngeal (type I), hypopharyngeal (type II), or both (type III). Since the severity of OSA as determined by PSG does not correlate well with surgical outcome,16,17,18,19,20 Friedman developed a staging system based on tonguepalate position, tonsil size and body mass index (BMI), that indicates which patients should be better surgical responders.21,22,23	We do not address the clinical practice decisions for how physicians choose which intervention, including which surgery, to use. This is beyond the scope of this CER.

Reviewer Name[1]	Reviewer Affiliation[2]	Section[3]	Reviewer Comments	Author Response4
Kim Brew	Medtronic	What is the comparative effect of different treatments for obstructive sleep apnea in adults? • Does the comparative effect of treatments vary based on presenting patient characteristics, severity of obstructive sleep apnea, or other pretreatment factors? • Are any of these characteristics or factors predictive of treatment success? • Characteristics: Age, sex, race, weight, bed partner, airway, other physical characteristics, and specific comorbidities • Obstructive sleep apnea	As mentioned in response to KQ1 the discussion did not address the physical characteristics of the airway. It also did not mention any diagnostic evaluation of the airway. By not identifying the site of anatomical obstruction, various therapies may or may not be actually affecting the cause of the patient's OSA. The surgical interventions reviewed in the document were limited to UPPP, laserassisted uvuloplasty (LAUP), radiofrequency ablation (RFA) of the tongue base, mandibular osteotomy with genioglossal advancement (GA), nasal septoplasty and radiofrequency inferior turbinate reduction. This is an incomplete list of surgical interventions for OSA. Historically, several operative procedures have been developed since the early 1980's to enlarge the pharyngeal airway and consequently prevent its obstruction during sleep.24,25,26,27,28 Initially, airway compromise was thought to be primarily in the retropalatal region, which led to operative procedures to expand the oropharyngeal area, most commonly UPPP with tonsillectomy if tonsils are present.29,30 More recently, a procedure to stiffen the soft palate have shown promise in preventing oropharngeal airway collapse.31,32 Because of the ineffectiveness of UPPP in a substantial number of patients, attention soon shifted to the retrolingual region as another site of airway compromise, which led to surgical procedures to expand the hypopharyngeal area or to keep the tongue from falling backwards into the hypopharynx during sleep, such as tongue base reduction, hyoid suspension, mandibular osteotomy with genioglossus advancement, and tongue base suspension.15,24,25,26,27,33,34,35,36 Where both oropharyngeal and hypopharyngeal compromise are unable to be corrected by soft 4 tissue ablation and repositioning procedures, especially in persons with facial skeletal deformities such as retrognathism, mandibular	We have added language to clarify that only interventions (which includes specific surgeries) that were evaluated by studies that met eligibility criteria were reviewed. The CER is not a narrative review or text book on the management of OSA. We were limited by the published evidence. We recognize that surgical studies are at a disadvantage since so few surgical RCTs have been done. This is why we made an exception for surgical studies and included large cohort studies.

Reviewer Name[1]	Reviewer Affiliation[2]	Section[3]	Reviewer Comments	Author Response4
Kim Brew	Medtronic	Conclusion	Medtronic appreciates the opportunity to provide comments on this CER. We recognize the need for well-developed clinical studies and a strong interest from all of the stakeholders in the field of sleep medicine to develop these studies. Due to the significant amount of influence that CERs may have on consumer, provider and payor decisions regarding the diagnosis and treatment of a medical condition, we recommend that current diagnostic tests and surgical interventions for OSA be fully represented in this CER. Again, we recognize the need for well-developed clinical studies and a strong interest from all of the stakeholders in the field of sleep medicine to develop these studies.	We hope our CER acts as a catalyst for more and better studies to address many of the gaps in the evidence.

Source: www.effectivehealthcare.ahrq.gov
Published Online: October 25, 2011

Reviewer Name[1]	Reviewer Affiliation[2]	Section[3]	Reviewer Comments	Author Response4
Tim Chestnut		How do different available tests compare to diagnose sleep apnea in adults with symptoms suggestive of disordered sleep? a. How do the different tests compare in different subgroups of patients, based on: race, gender, body mass index (BMI), existing non-insulin dependent diabetes mellitus (NIDDM), existing cardiovascular disease (CVD), existing hypertension (HTN), clinical symptoms, previous stroke, airway characteristics b. What is the	An extensive review of this question was undertaken by an AHRQ technology assessment program and published in 2007.The emphasis focused on evaluation of portable vs in-lab reporting of AHI as a predictor of treatment in patients with OSAHS. We know that AHI alone correlates poorly with symptoms, which means that AHI alone is not sufficient for diagnosis. AHI also correlates poorly with treatment response and adherence. Therefore AHI identifies patients 'suggestive' of OSAHS, not diagnostic of the syndrome alone. Measurement of AHI with either in-lab studies, or portable monitors of types II-IV adequately measured AHI. The classification system is outmoded and is in need of revision; it does not adequately categorize peripheral arterial tonometry. However portable monitors incorporating this technology were reviewed as well, and found to acceptably measure AHI. It should be noted that several groups are currently employing the use of portable monitors for diagnosis and management of OSAHS patients. These include Kaiser/Group Health in California, Oregon and Washington, as well as the VA,(Sleep Diagnosis and Therapy v.2: 21-23, Nov 2007). Many other groups are using home monitors in similar programs and we have recently completed a pilot study with primary care practitioners. In addition, I am aware of other research projects currently in progress evaluating similar models. a. Most data available describing risks of having OSAHS have relied on studies with high risk groups predominantly males with a mean age of about 50 years, and a BMI of over 25kg/m¬2. Although there is increasing epidemiologic data available examining the risks in subgroups, very little data is available regarding the application of alternative diagnostic strategies to these groups. Considerable data is available on the evaluation of sleep disordered breathing in patients with cardiac disease, including complex sleep apnea and Cheyne-Stokes breathing. Most of the studies validating the use of portable monitors included a	We have improved our descriptions of the definitions of OSA. The report has no opinion on how the definitions should be changed. The report is not meant to address the current use of any of the studied interventions. We believe we have fully addressed all the Key Questions based on the literature that met agreed upon eligibility criteria. We agree that many other research questions are of interest.

Reviewer Name[1]	Reviewer Affiliation[2]	Section[3]	Reviewer Comments	Author Response4
Tim Chestnut		How does phased testing (screening tests or battery followed by full test) compare to full testing alone?	Full testing 'alone' provides an AHI suggestive of OSAHS but not diagnostic of the disorder. Detection of OSAHS must rely on history, symptoms and clinical signs. A formal clinical prediction rule may then be calculated to allow the selection of patients who are appropriate to undergo formal testing. The combination then allows the diagnosis of OSAHS. Selection of the formal test was covered in key question 1.	This appears to be the reviewer's answer to the question. No change made to the report.
Tim Chestnut		What is the effect of pre-operative screening for sleep apnea on surgical outcomes?	Clarification is necessary for this question. Two applicable areas may be relevant. First, preoperative screening for patients undergoing anesthesia has been recommended to reduce the postoperative risks. The American Society of Anesthesiologists has published a guideline in Anesthesiology in May of 2006, (v.104:1081-93), covering this issue. No formal research evaluation of this effectiveness has been published. Chung has published another, more recent review,(Curr. Opion. Anesthesiol. 22:405-11, 2009). The second area that may be relevant to this question is the use of correct techniques to evaluate patients who are potential surgical candidates for treatment of OSAHA. This is considered in two circumstances; when there is a correctable upper airway anatomic abnormality, and when PAP is either intolerable or ineffective.	This appears to be the reviewer's answer to the question. No change made to the report.

Reviewer Name[1]	Reviewer Affiliation[2]	Section[3]	Reviewer Comments	Author Response4
Tim Chestnut		What is the comparative effect of different treatments for obstructive sleep apnea (OSA) in adults? a. Does the comparative effect of treatments vary based on presenting patient characteristics, severity of OSA, or other pre-treatment factors? Are any of these characteristics or factors predictive of treatment success? Characteristics: Age, sex, race, weight, bed partner, airway and other physical characteristics, specific comorbidities OSA severity or characteristics: Baseline questionnaire	This is a very complex and important question which remains largely unanswered despite the progress in the last decade. Wright and colleagues published a review in 1997, (BMJ 314:851-8660), outlining the scarcity of outcome and effectiveness data regarding treatment of OSA. Since that time the literature has grown steadily. Recent reviews include Schwartz et al,(Proc Am Thor Soc,5:185-92,2008), and Punjabi,(Proc Am Thor Soc,5:136-43,2008). The use of CPAP, and alternative therapies was recently reviewed and published by the American Thoracic Society as a Symposium in 2008. Sanders et al reviewed the use of CPAP, (Proc Am Thor Soc 5:161-172,2008), and the issue of titration in lab or at home, by technicians vs. autotitrating devices. In moderate to severe patient groups the outcomes were essentially the same. Other more modest and mild severity group, (as defined by RDI) ,need further study. The AASM has published guidelines for the use of autotitrating devices,(Sleep 31:141-147,2008). In addition this issue was addressed in the 2007 AHRQ study. See the comments in question 5, below. b. Does the comparative effect of treatments vary based on the definitions of OSA used by study investigators? This is an important and unanswered question. A large percentage of patients with OSAHS have mild or moderate degrees of disordered breathing. Given the prevalence of sleep disordered breathing heterogeneous group, it will be a very difficult question to answer. We do know that an elevated risk of morbidity and mortality extends to this milder group of patients. Some work has been done, but this is an area urgently in need of further clinical research.	Reviewer does not suggest any revisions to report.

Reviewer Name[1]	Reviewer Affiliation[2]	Section[3]	Reviewer Comments	Author Response4
Tim Chestnut		What are the pre-treatment patient-level characteristics associated with failure to adequately treat OSA?	The prediction of success and failure in treating OSAHS is a complex issue. The major issue in all groups was long term adherence, an issue reviewed in detail by Weaver and colleagues in the same symposium, (Proc Am Thor Soc 5:173-8 ,2008). Outcome measures and the best way to judge adequacy of treatment have recently been reviewed by Sanders and Weaver in the ATS symposium. However there is clearly a lack of good data to judge which outcome measures are predictors of subsequent morbidity and mortality, so adequacy is difficult to judge. Given that the study of pre-treatment characteristics will be difficult to ascertain. Two recent extensive literature reviews were published by Olsen et al, (Clin Psychol Rev 28:1355-1371, 2008), and Poulet et al, (Sleep Med. 10:993-99,2009). Although some predictors of success were outlined, much work remains to be done comparing the effectiveness of various interventions to improve adherence.	Reviewer does not suggest any revisions to report.
Tim Chestnut		What is the effect of interventions to improve compliance with device (CPAP, oral appliances, positional therapy) use on clinical and intermediate outcomes?	The predictors of CPAP success have become increasingly important since the poor adherence has been recognized and the benefits of CPAP use have become clearer. Weaver et al recently published an excellent review of predictors of adherence, and interventions to improve the long term use of therapy. Most of the literature covering improved compliance overlaps with the reviews covered under question 5, above. One area of recent notable advance is the ability with home autotitrators of performing remote monitoring of compliance, and area of active research at this time.	Reviewer does not suggest any revisions to report.

Source: www.effectivehealthcare.ahrq.gov
Published Online: October 25, 2011

Reviewer Name[1]	Reviewer Affiliation[2]	Section[3]	Reviewer Comments	Author Response4
Terence Davidson		How do different available tests compare to diagnose sleep apnea in adults with symptoms suggestive of disordered sleep? a. How do the different tests compare in different subgroups of patients, based on: race, gender, body mass index (BMI), existing non-insulin dependent diabetes mellitus (NIDDM), existing cardiovascular disease (CVD), existing hypertension (HTN), clinical symptoms, previous stroke, airway characteristics b. What is the	1. The simplest and most effective paradigm is clinical suspicion, based on history and examination followed by measurement of the apnea hypopnea index by Types I, II, II, III or IV sleep studies. Treatment recommendations are based primarily on the AHI using Medicare guidelines or clinical suspicion in patients whose AHI does not fulfill current Medicare guidelines. There are relatively few who fall into this group. The primary individuals are young women who probably have upper airway resistance syndrome but are more easily diagnosed by PAP therapy trial than by PSG with esophageal balloons. a. Other than those with Central Sleep Apnea, predominantly those with Congestive Heart Failure or central nervous system disease, patients are diagnosed and treated as noted above. 2. Diagnosis is based on the combination of the above. Given the morbidity and mortality of obstructive sleep apnea, those highly suspect for OSA are and should be given a PAP therapy trial.	Reviewer does not suggest any revisions to report.

Reviewer Name[1]	Reviewer Affiliation[2]	Section[3]	Reviewer Comments	Author Response4
Terence Davidson		How does phased testing (screening tests or battery followed by full test) compare to full testing alone?	1. Phase testing is basically a non-productive use of time and money. Those suspect for OSA need a sleep test and a PAP therapy trial.	Reviewer does not suggest any revisions to report.
Terence Davidson		What is the effect of pre-operative screening for sleep apnea on surgical outcomes?	1. Neither history nor sleep test predicts surgical outcomes. Anatomy is the sole predictor of surgical outcomes. Friedman's Staging Symptom is a good predictor for tonsillectomy, nasal surgeries for total nasal obstruction are sometimes effective in treating sleep apneas but are indicated based on their nasal obstruction including nasal polyps for chronic sinusitis and anatomic obstruction, be it deviated septum, nasal fracture and deformity or valve collapse.	Reviewer does not suggest any revisions to report.

Source: www.effectivehealthcare.ahrq.gov
Published Online: October 25, 2011

Reviewer Name[1]	Reviewer Affiliation[2]	Section[3]	Reviewer Comments	Author Response4
Terence Davidson		What is the comparative effect of different treatments for obstructive sleep apnea (OSA) in adults? a. Does the comparative effect of treatments vary based on presenting patient characteristics, severity of OSA, or other pre-treatment factors? Are any of these characteristics or factors predictive of treatment success? Characteristics: Age, sex, race, weight, bed partner, airway and other physical characteristics, specific comorbidities OSA severity or characteristics: Baseline questionnaire	1. Currently these are clinical decisions based on history, physical exam and the apnea hypopnea index. Generally speaking, the worse the disease, the heavier the patient and the worse the anatomy, the more likely that PAP therapy is the only treatment. Patients with retrognathia who fail PAP therapy are generally recommended for maxillomandibular advancements or in mild cases, mandibular advancement devices. Surgical recommendations and therapies are based on anatomy—Friedman's Staging System being as good as any. 2. Implied but not directly asked is the value of screeners. Attached is a recent metaanalysis of screeners. The better screeners have a 90% sensitivity, 70% specificity and they seem best applied in the anesthesia pre-operating clinic and perhaps for mass secreening for large groups, such as truck and bus drivers, albeit honesty in these groups is questionable. The bottom line is that the Apnea-Hypopnea Index (AHI) is the best objective measure of OSA sensitivity. Home sleep testing measures the AHI just as well as polysomnography (PSG) and provides the number at a quarter of the cost of PSG with significantly improved patient acceptance. We need to move our energies into diagnosing more people and improving PAP therapy compliance. We need to stop focusing our attention on preserving the expensive PSG.	The attached meta-analysis did not provide any studies that met our criteria to address the Key Questions evaluated for this report. This report has no opinion on what policymakers ought to do.

Source: www.effectivehealthcare.ahrq.gov
Published Online: October 25, 2011

Reviewer Name[1]	Reviewer Affiliation[2]	Section[3]	Reviewer Comments	Author Response4
Jon Freudman		How do different available tests compare to diagnose sleep apnea in adults with symptoms suggestive of disordered sleep? a. How do the different tests compare in different subgroups of patients, based on: race, gender, body mass index (BMI), existing non-insulin dependent diabetes mellitus (NIDDM), existing cardiovascular disease (CVD), existing hypertension (HTN), clinical symptoms, previous stroke, airway characteristics b. What is the	AHRQ published a review of home sleep testing for OSA in August 2007 . A large body of evidence was considered including 22 studies involving type III devices. The following conclusion was reached: Type III home sleep testing "may identify AHI suggestive of OSAHS with high positive likelihood ratios and low negative likelihood ratios". In the years since this analysis there has been a concerted effort by those with a vested interested in maintaining the high volume of expensive testing (costs range from $1,000 to $3,500 per test) at sleep labs to refute the reliability of home sleep testing. Those performing the current AHRQ comparative effectiveness evaluation should be cautious regarding manipulation of this new review by sleep medicine specialists who are not economically neutral on this issue.	The new evidence did not substantively change the conclusions of our 2007 report.

Reviewer Name[1]	Reviewer Affiliation[2]	Section[3]	Reviewer Comments	Author Response4
Jon Freudman			2) When considering the diagnostic comparative effectiveness of Home Monitoring Devices (portable monitoring-PM) vs. Polysomnography (PSG) it is important to evaluate performance in both arms, not just home monitoring. For example, the study of Berry et al evaluated patients randomized to PM or PSG. Of the 53 patients randomized to testing by PM, 4 patients were without OSA. One of these was found to have OSA when subsequently retested by PSG. Of the 53 patients randomized to testing by PSG, 6 were without OSA. Two of these were found to have OSA when retested by PM. Thus the excellent parallel design of this study allows for a meaningful comparison. In contrast, the AASM is sponsoring a study that alleges to compare PSG with home monitoring . In this study patients in the home monitoring arm who test negative for OSA are retested utilizing PSG. However, those patients in the PSG arm who test negative for OSA are not retested by portable monitoring or PSG. Thus by design, this AASM sponsored study will detect patients missed by PM but not those missed by PSG.	We agree that verification bias is a significant source of bias in diagnostic studies, and it is very difficult to adjust for confounding once verification bias is present. For this reason we excluded studies we thought were at high risk of verification bias. The Berry study was excluded as it did evaluated clinical pathways (including treatment) which were not directly relevant to the Key Question. In addition, this study had verification bias as only those screened out for OSA were retested with PSG.

Reviewer Name[1]	Reviewer Affiliation[2]	Section[3]	Reviewer Comments	Author Response4
Jon Freudman			3) When comparing PM and PSG it is important to consider the fact that that there is night-to-night variation in OSA and that difficulty sleeping in a sleep lab (the “first night” effect) results in many non-diagnostic tests. There is a literature substantiating that due to night-to-night variation and the “first night effect” (FNE) one night of PSG is insufficient for the diagnosis of OSA for approximately 15-25% of cases . Because of the complexity of the PSG testing process, the manual analysis used by PSG labs to score the PSG data, and the first night effect, in a significant number of PSGs a definitive diagnosis of OSA cannot be achieved. Home sleep studies allow for testing in the patient’s usual sleep environment as well as for several nights of testing; consequently, this eliminates the first night effect and captures cases that might be missed because of night-to-night variation inherent in the PSG testing process. Therefore, when considering the comparative effectiveness of PM vs. PSG, the first night effect, night-to-night variation, and the need to repeat PSGs should be factored into the equation.	The studies did not allow us to adequately assess any issues related to first night effect or night to night variation. We take note of the comment, and have included it in our discussion section (page 138, near bottom).

Reviewer Name[1]	Reviewer Affiliation[2]	Section[3]	Reviewer Comments	Author Response4
Jon Freudman			4) When considering the trade offs between PM and PSG, the economic benefits of earlier diagnosis need to be modeled. The AHRQ analysis of December 4, 2007, Obstructive Sleep Apnea-Hypopnea Syndrome: modeling different diagnostic strategies, concluded that PM strategies would result in the initiation of OSA treatment on average 6 months earlier than sleep lab strategies. However this analysis did not attempt to model the economic advantages of this earlier diagnosis and treatment of OSA. Given that there is randomized control trial evidence regarding the quality of life improvements and antihypertensive benefits of CPAP, there can no longer be ethical RCTs comparing CPAP with no treatment that last longer than a few weeks. Thus, when considering the benefits of CPAP and the benefits of earlier initiation of CPAP, modeling will be required and should include costs of physician office visits, clinical lab tests, pharmaceutical usage, durable medical equipment, hospital and ER utilization, QUALYS, and worker productivity/lost time at work/absenteeism. The association between untreated OSA and motor vehicle accidents is well described as is the beneficial impact of CPAP on the reduction of accidents in these patients , . If a PM strategy is associated with initiation of CPAP 6 months earlier than a PSG strategy, the impact on reduction of motor vehicle accidents should be modeled as part of a comparative effectiveness analysis. Elimination of motor vehicle accidents will have enormous economic impact. Similarly, the impact of a 6-month earlier diagnosis of OSA on the cost of healthcare should be modeled. Multiple studies have shown that people with untreated OSA have markedly increased healthcare costs and that the diagnosis and treatment of OSA offsets this increase. Albarrack et al showed that treatment of OSA reversed the trend of healthcare utilization seen prior to diagnosis. Brahamman et al documented a reduction in professional claims	We agree this is an interesting and important question, but it is beyond the scope of this CER.

Reviewer Name[1]	Reviewer Affiliation[2]	Section[3]	Reviewer Comments	Author Response4
Jon Freudman			5) When considering the comparative effectiveness of PM vs PSG, the evidence in support of auto-titrating CPAP needs to be considered. Auto-titration is an integral aspect of a PM strategy. Auto CPAP can be used to replace the initial PSG CPAP titration study conducted in a sleep lab (currently accounts for 40% of the total sleep study utilization in the US) or chronic use of autoPAP would eliminate the need for a repeat PSG titration study (usually required if there is a change in the OSA patient's weight by 10% up or down, a change in the co-morbid situation, or a change in the risk factors associated with that patient) and the evidence in support of this is discussed under key question 4. Auto-titrating CPAP can also be utilized on a short-term (i.e. multi-night unattended home titration study) basis for home titration studies thus obviating the need for sleep lab titration. This strategy has been widely utilized at the Veterans Administration and at Kaiser Permanente. The literature in support of this strategy is extensive and includes randomized controlled trials. Of note is that all the studies have showed comparable outcomes (Epworth Sleep Scale, AHI) with PM and PSG. Portable monitoring and autoPAP was associated with cost savings in the Rice study and improved adherence to therapy in the Mulgrew randomized controlled trial. AHRQ Home diagnosis of Obstructive Sleep Apnea-Hypopnea Syndrome August 8, 2007 Berry RB; Hill G; Thompson L; McLaurin V. Portable monitoring and autotitration versus polysomnography for the diagnosis and treatment of sleep apnea. SLEEP 2008;31(10):1423-1431. Portable monitoring for diagnosis and management of sleep apnea (HomePAP). Available at clinicaltrials.gov/ct2/show/NCT00642486. Ahmadi N, Shapiro G, Chung S and Shapiro C. Clinical diagnosis of sleep apnea based on single night of polysomnography vs. two nights of polysomnography Sleep Breath December 2008	This CER did not evaluate titration, including the need for PSG to perform titration or different methods to find an appropriate pressure. It was not a Key Question.

Reviewer Name[1]	Reviewer Affiliation[2]	Section[3]	Reviewer Comments	Author Response4
Jon Freudman		What is the comparative effect of different treatments for obstructive sleep apnea (OSA) in adults? a. Does the comparative effect of treatments vary based on presenting patient characteristics, severity of OSA, or other pre-treatment factors? Are any of these characteristics or factors predictive of treatment success? Characteristics: Age, sex, race, weight, bed partner, airway and other physical characteristics, specific comorbidities OSA severity or characteristics: Baseline questionnaire	autoPAP In addition to the literature regarding auto-titrating CPAP for titration studies, there is also high-level evidence documenting the efficacy and advantages of autoPAP for chronic therapy. In randomized controlled trial comparing fixed autoPAP, Massie et al documented equivalent improvement in daytime sleepiness (ESS) and better patient acceptance of auto-PAP as measured by SF Vitality and SF 36 Mental health scores. In this study autoPAP was also associated with more restful sleep, overall better quality of sleep, less discomfort from pressure when compared with fixed does CPAP. In another randomized controlled trial comparing fixed and auto-titrating CPAP, Hukins reported equivalency in ESS improvement. However auto-titrating CPAP was associated with fewer reported side effects and air leaks. Planes et al compared fixed CPAP and autoPAP in a randomized controlled trial involving patients with severe OSA. The modalities were equivalent in reduction of AHI, ESS and improvements in sleep quality. AutoPap was associated with prompter initiation of therapy and significant cost savings. Nussbarmer et al performed a randomized double blind, controlled cross over trial comparing fixed CPAP and autoPAP. Improvements in ESS and AHI were equivalent however there was a statistically significant patient preference for autoPAP. The literature is clear that autoPAP is as effective as fixed pressure CPAP and patients prefer that autoPAP. Indeed a recent Cochrane Database Systematic Review evaluated 45 randomized controlled trials (1874 participants) and confirmed equivalent ESS outcomes and patient preference for autoPAP. Given that the CPAP compliance literature has documented that a patient's initial experience with CPAP impacts later compliance, early patient preference for autoPAP is extremely important. Massie C et al. Comparison between Automatic and Fixed Positive Airway Pressure Therapy in the HomeAm J Respir Crit Care Med Vol 167. pp 20–23, 2003 Hukins Comparative	This appears to be the reviewer's answer to the question. No change made to the report.

Reviewer Name[1]	Reviewer Affiliation[2]	Section[3]	Reviewer Comments	Author Response4
Jon Freudman		What is the effect of interventions to improve compliance with device (CPAP, oral appliances, positional therapy) use on clinical and intermediate outcomes?	See discussion of autoPAP in question 4. RCTs have confirmed patient preference for autoPAP vs fixed CPAP	This appears to be the reviewer's answer to the question. No change made to the report.

Source: www.effectivehealthcare.ahrq.gov
Published Online: October 25, 2011

Reviewer Name[1]	Reviewer Affiliation[2]	Section[3]	Reviewer Comments	Author Response4
David Nielsen	AAO-HNS (American Academy of Otolaryngology – Head and Neck Surgery)		While we generally agree that the OSA CER is fair and reasonably balanced we urge AHRQ to include a clear acknowledgement that there are inherent challenges for surgical treatments to obtaining the level of evidence you have chose to accept in this review. As stated in the review, compliance with Continuous Positive Airway Pressure (CPAP) treatment is poor. Furthermore, surgical treatment is often used in cases of failure of CPAP therapy and in those patients, is the only remaining treatment option available. While the heterogeneity in the surgery studies precludes a general conclusion, the individual studies support an important effect of surgery. We assert that the effect of surgical treatment on survival appears, in the limited studies available, to be equal or superior to that of CPAP (references 241 and 244). There is no evidence to say surgery is less effective on clinical outcomes compared to the other main treatments. When patients are unable to use the other treatments, surgery is the only remaining option and absolutely should be considered.	We have described the individual findings of the 12 relevant studies for all outcomes, including the 2 retrospective analyses of survival. We do not claim any evidence to support that surgery is less effective on clinical outcomes compared to the other main treatments. Due to the limitations of the evidence, the overall conclusion remains that there is insufficient evidence regarding this comparison of treatment options.
Dov Rubin		Chapter 4. Summary & Discussion	Focus on outcomes seems to neglect and overlook a large body of evidence including the Blue Cross evaluation by the California Technology Assessment Forum (CTAF) in which it only found 2 diagnostic OSA devices worthy of changing outcomes, with over 200 diagnostics with a PSG, and full home evaluation. http://www.ctaf.org/content/assessment/detail/1038	This report is not a review of other groups' reviews or assessments of the literature. We have no comment on their report.

Reviewer Name[1]	Reviewer Affiliation[2]	Section[3]	Reviewer Comments	Author Response4
Dov Rubin		References	The following references are missing as part of evidence based outcomes for ambulatory devices: Blue Cross CTAF: http://www.ctaf.org/content/assessment/detail/1038 Richard B. Berry, Gilbert Hill, Linda Thompson, Valorea McLaurin, Malcom Randall. Portable Monitoring and Autotitration versus Polysomnography for the Diagnosis and Treatment of Sleep Apnea. SLEEP 2008; 31(10):1423-1431. Choi JH, Kim EJ, Kim YS, Choi J, Kim TH, Kwon SY, Lee HM, Lee SH, Shin C, Lee SH. Validation study of portable device for the diagnosis of obstructive sleep apnea according to the new AASM scoring criteria: Watch-PAT 100. Acta Otolaryngologica 2010; 130(7):838-43 Pillar G, Bar A, Bettito M, Schnall R, Dvir I, Sheffy J, Lavie P. An automatic ambulatory device for detection of AASM defined arousalsfrom sleep: the WP100. Sleep Medicine2003; 4(3):207-212. Bar A, Pillar G, Dvir I, Sheffy J, Schnall RP, Lavie P. Evaluation of a Portable Device Based on Peripheral Arterial Tone for Unattended Home Sleep Studies. Chest 2003; 123(3): 695-703. Ayas N. TA, Pittman S, MacDonald M, White D. Assessment of a Wrist-worn Device in the Detection of Obstructive Sleep Apnea. Sleep Medicine 2003; 4(5):435-442. Pittman DS, Ayas NT, MacDonald MM, Malhotra A, Fogel RB, White D. Using a Wrist-Worn Device Based on Peripheral Arterial Tonometry to Diagnose Obstructive Sleep Apnea: In-Laboratory and Ambulatory Validation. SLEEP 2004; 27(5):923-933. Hedner J, Pillar G, Pittman DS, Zou D, Grote L, White D. A Novel adaptive wrist actigraphy algorithm for Sleep-Wake assessment in sleep apnea patients. SLEEP 2004; 27(8):1560-1566. Zou D, Grote L, Peker Y, Lindblad U, Hedner J. Validation a Portable Monitoring Device for Sleep Apnea Diagnosis in a Population Base Cohort Using Synchronized Home Polysomnography. SLEEP 2006; 29(3):367-374. Pittman S.D, Pillar G, Berry RB, Malhotra A,	These studies were reviewed. None met eligibility criteria based on intervention, design, or reported analyses.

Reviewer Name[1]	Reviewer Affiliation[2]	Section[3]	Reviewer Comments	Author Response4
Dov Rubin		Tables and Figures	Comment on Table A: The pre-occupation with Device Types (I, II, III, IV) seems overly exaggerated. Focus should be on proven evidence-based outcomes.	We used device types to organize the interventions. Regardless of the organization we evaluated all appropriate (evidence based) outcomes.
Patrick Strollo Jr	AASM (American Academy of Sleep Medicine)		We feel that the value of the apnea-hypopnea index (AHI) should bot be discounted or downgraded as a measure of the effectiveness of treatments of obstructive sleep apneas (OSA). Since OSA is, by definition, a sleep-related breathing disorder, the resolution of apneas and hypopneas during sleep constitutes successful treatment. We believe that there is some confounding between the questions of "what treatments effectively eliminate OSA?" versus "how important is it to treat OSA?" The latter would include the relationship between AHI and clinical outcomes such as all-cause mortality, cardiovascular mortality and quality of life. This philosophy was used to downgrade evidence from high to moderate for continious positive airway pressure (CPAP) devices. The level of evidence was given as equivalent between CPAP and mandibular advancement devices (MADs) even though there were 5 times as many studies on CPAP than on MAD. In addition, adherence to MADs has not been objectively measured as is inaccurate compared to objective measures. Our reviews rate the evidence supporting the effectiveness of CPAP to treat OSA as high. This is further supported in the report of the direct comparison of CPAP to MAD, which states that there is a moderate level of evidence that CPAP is more effective than MAD.	It appears that the reviewer has misunderstood our methods. The evidence on CPAP was initially downgraded due to lack of clinical outcomes reported. Using the AHRQ approach, this is completely appropriate regardless of any association between intermediate and clinical outcomes. Subsequently, we "broke the rules" somewhat and upgraded the evidence because of the very large effect of CPAP on AHI. Note that we are most interested in the effect of treatment on clinical outcomes (including symptoms) not on laboratory measurements.

Source: www.effectivehealthcare.ahrq.gov
Published Online: October 25, 2011

Reviewer Name[1]	Reviewer Affiliation[2]	Section[3]	Reviewer Comments	Author Response4
Patrick Strollo Jr		Key question #2 (phased assessment)	We feel that the studies of Mulgrew (2007) and Berry (2008) should be included in the assessment for Key Question #2 (phased assessment). We realize there were reasons for their exclusion, but these reasons seem more applicable to exclusions for Key Question #1 (methods for diagnosis).	Key Question #2 is aimed at evaluating phased testing (a series of tests performed dependent on the results of initial tests) with full testing (overnight PSG) alone. In both Mulgrew and Berry, the research question addressed comparisons made between different clinical pathways (including diagnosis and treatment) rather than a comparison of phased testing with PSG. In Mulgrew 2007, the clinical instruments were sequentially utilized, primarily to ensure that only participants with a high pre-test probability of OSA were recruited. Comparisons with PSG results were not made. In Berry 2008, the comparisons involved 2 different clinical pathways: portable monitoring for diagnosis and treatment with unattended autoCPAP with PSG for diagnosis and treatment of OSA. While Both studies answered clinically relevant questions, the comparisons in the studies were not relevant to our key question.
Patrick Strollo Jr			It should be acknowledged that virtually all the studies that examined diagnosing and treating OSA in the absence of an attended polysomnogram were conducted y sleep physician within centers of excellence. Whether out of laboratory approaches utilizing portable cardiopulmonary monitors and auto-titrating CPAP can achieve similar acceptance and adherence and subsequent outcomes in the absence of input from sleep specialist and/or center remains unstudied. Prior data reported in the Journal of Clinical Sleep Medicine 2006 2: 133-42 by Parthasarthy et al suggests that this is not the case. Further work in this area is vital for defining value based disease management pathways.	This is an interesting question, but largely beyond the purview of this report. We have added a paragraph in the Discussion (page 140). Questioning the generalizability of the academic / research setting of the studies to the general population.

Source: www.effectivehealthcare.ahrq.gov
Published Online: October 25, 2011

Reviewer Name[1]	Reviewer Affiliation[2]	Section[3]	Reviewer Comments	Author Response4
Sharon Tracy		Chapter 3 Results	I think you need to reassess the effect of CPAP and autoPAP on quality of life. Seven studies were included in the assessment of autoPAP vs. CPAP with respect to quality of life, with the result that there are no significant differences between autoPAP and CPAP on quality of life. However, these studies were not included in the assessment of the effect of either CPAP or autoPAP on quality of life. The inclusion of these data may make a meta-analysis possible and also show a statistically significant improvement in quality of life with positive airway pressure. This reassessment may change the level of evidence from moderate to high in favor of PAP therapy. The articles are: Meurice 2007, Hukins 2004, Massie 2003, To 2008, Nussbaumer 2006, Senn 2003 and Fietze 2007. In any case, the data should be included for the sake of completeness.	As per our methods, agreed upon with the TEP, we have analyzed only direct comparisons. None of the studies that compared autoCPAP to CPAP were eligible for the evaluations of CPAP vs control (or sham). These studies would have to be treated as single arm cohort studies, which were not included in this report.

Reviewer Name[1]	Reviewer Affiliation[2]	Section[3]	Reviewer Comments	Author Response4
Pell Wardrop	**Chair Sleep Disorders Committee AAO=HNS**	What is the comparative effect of different treatments for obstructive sleep apnea (OSA) in adults? a. Does the comparative effect of treatments vary based on presenting patient characteristics, severity of OSA, or other pre-treatment factors? Are any of these characteristics or factors predictive of treatment success? Characteristics: Age, sex, race, weight, bed partner, airway and other physical characteristics, specific comorbidities OSA severity or characteristics: Baseline questionnaire	Peer-reviewed medical literature supports the use of CPAP for the treatment of obstructive sleep apnea . Standard CPAP, and the variants BiPAP and Auto-CPAP, have been found in randomized controlled trials to be highly efficacious . However, There is a critical distinction between efficacy and effectiveness (REFERENCE: Flay BR. Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs. Prev Med 1986;15(5):451-74.). Efficacy is the effect in the lab or under ideal circumstance, regardless of treatment compliance. CPAP demonstrates very good efficacy. Effectiveness is the effect in everyday life, which depends in part on patient compliance with therapy. CPAP demonstrates fair to good effectiveness. This distinction is critical, because surgical treatment effectiveness does not depend on compliance. Thus surgical treatment, while less efficacious than CPAP, can be as effective or more effective than CPAP (REFERENCES: 1. Peker Y, Hedner J, Norum J, Kraiczi H, Carlson J. Increased Incidence of Cardiovascular Disease in Middle-aged Men with Obstructive Sleep Apnea: A 7-Year Follow-up. Am J Respir Crit Care Med 2002;166(2):159-65. 2. Weaver EM, Maynard C, Yueh B. Survival of veterans with sleep apnea: continuous positive airway pressure versus surgery. Otolaryngol Head Neck Surg 2004;130(6):659-65. 3. Woodson BT, Steward DL, Weaver EM, Javaheri S. A randomized trial of temperature-controlled radiofrequency, continuous positive airway pressure, and placebo for obstructive sleep apnea syndrome. Otolaryngol Head Neck Surg 2003;128(6):848-61.). Because CPAP is low risk and because CPAP can be highly effective in some patients, it is appropriate first line therapy. Successful long term treatment of OSAS with CPAP is as defined by 4 hours of use 70% of nights. This is the equivalent of a minimum 2.8 hours of CPAP use nightly. By definition, successful CPAP treatment can mean effective treatment for as little as 35%	We do not believe that we have conflated or confused efficacy and effectiveness. We do not believe we can compare the efficacy or CPAP trials with the effectiveness of surgical trials. We included only direct comparisons and generally avoided indirect cross-study comparisons of interventions.

Reviewer Name[1]	Reviewer Affiliation[2]	Section[3]	Reviewer Comments	Author Response4

[1] Names are alphabetized by last name. Those who did not disclose name are labeled "Anonymous Reviewer 1," "Anonymous Reviewer 2," etc.
[2] Affiliation is labeled "NA" for those who did not disclose affiliation.
[3] If listed, page number, line number, or section refers to the draft report.
[4] If listed, page number, line number, or section refers to the final report.

Source: www.effectivehealthcare.ahrq.gov
Published Online: October 25, 2011