Comparative Effectiveness Review Disposition of Comments Report

Research Review Title: *Antipsychotics for the Prevention and Treatment of Delirium*

Draft review available for public comment from March 4, 2019 to April 1, 2019.

Research Review Citation: Neufeld KJ, Needham DM, Oh ES, Wilson LM, Nikooie R, Zhang A, Koneru M, Balagani A, Singu S, Aldabain L, Robinson KA. Antipsychotics for the Prevention and Treatment of Delirium. Comparative Effectiveness Review No. 219. (Prepared by the Johns Hopkins University Evidence-based Practice Center under Contract No. 290-2015-00006-I-2.) AHRQ Publication No. 19-EHC019-EF. Rockville, MD: Agency for Healthcare Research and Quality; September 2019. Posted final reports are located on the Effective Health Care Program search page.
DOI: https://doi.org/10.23970/AHRQEPCCER219.

Comments to Research Review

The Effective Health Care (EHC) Program encourages the public to participate in the development of its research projects. Each research review is posted to the EHC Program Web site or AHRQ Web site in draft form for public comment for a 3-4-week period. Comments can be submitted via the 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 research review.

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

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

Commentator & Affiliation	Section	Comment	Response
TEP Peer Reviewer #1	Methods	Was there a systematic approach to determining which analyses would get meta analyses? At times, the use of meta analyses seem arbitrary. For example, on page 37 – three studies (low, unclear, and high risk of bias), three different delirium detection instruments, and wide setting differences underwent meta analysis. Yet, length of stay eluded meta analysis despite a large number of studies and standardized measurement (despite the skew). (page 28). Could the authors please add more language to the methods section on the use of meta analysis.	Thank you for your comment. Decisions on what to pool in a meta-analysis can often be subjective. We clarified in the Data Synthesis section that outcome definition was a key variable for determining homogeneity and that we had discussed and decided as a team what could be pooled in a meta-analysis. We considered conducting a meta-analysis of length of stay in hospital. However, as we have detailed in the Results section, most of the studies reported median values. Meta-analysis of median values is generally not acceptable and we determined that transformation of the data was not appropriate due to their skewed nature. Ideally, the studies would have reported hazard ratios, which could have been pooled. We added a note about this in the Methods section.
TEP Peer Reviewer #1	Results	A technical point. The forest plots are labeled as meta analyses, technically they are forest plots. More important, the delirium studies are dealing with small effects (RR<2.0), yet the logarithmic scale seems a little out of proportion (.1 to 10). Some studies use arrows to represent wide confidence intervals (beyond the scale). From a readers standpoint, it is hard to differentiate RR=1.09 from RR=1.8. It is also difficult to determine when the pooled effect crosses 1.0, especially on the confidence intervals. A great example of this is in Figure 16, where the pooled RR is below 1, but the visualization appears to be above 1.0.	We have revised the scales on the forest plots to better fit the data.
TEP Peer Reviewer #1	Methods	Can the reviewers add methodology around the exclusion of studies from meta analysis to determine the contribution effect of each? It is a technique that is used frequently and not detailed.	We have added to the Methods, “We also conducted sensitivity analyses by omitting one study at time to assess the influence of any single study on the pooled estimates.”
TEP Peer Reviewer #1	Overall	In several areas there is a referral to the ‘overall’ section. However, this section is not labeled in the report. I suspect it means the parent question (rather than the subsection). Could this be cross-referenced and clarified	We have added the term “Overall” to the headings in the overall results section.
TEP Peer Reviewer #1	Results	Some sections add the strength of evidence in parentheses (page 36 KQ1a). This is helpful to me as a reader	Thank you for your comment. We graded only the outcomes that were considered critical by the Key Informants and Technical Experts. We have reviewed the report, and ensured that we have provided the strength of evidence in parentheses for those outcomes.

Source: https://effectivehealthcare.ahrq.gov/products/antipsychotics/research
Published Online: September 3, 2019

Commentator & Affiliation	Section	Comment	Response
TEP Peer Reviewer #1	Results	The combination of second generation antipsychotics into a single group is troubling. The receptor binding coefficients of these medications varies substantially (i.e. they are not all D2 antagonists). It is probably a predetermined outcome, but a statement of acknowledging the breadth of receptors touched would be appropriate.	We added an acknowledgement of the different mechanism of actions to the Methods chapter, Data Synthesis section. We now state, "Although the drugs may have different mechanisms of action, we anticipated that most drugs within a class would have similar clinical effects. Therefore, we combined studies of unique medications within classes when reporting outcomes." As noted in the methods, we explored the effects of pooling unique medications if there was substantial heterogeneity.
TEP Peer Reviewer #1	Executive Summary	For Figures A&B are the colors a AHRQ standard? a. The names and colors are duplicative (although this hits the points). b. Is the minus sign (negative effect) too similar to the arrow (no effect)?	The colors for Figures A and B were selected to maximize contrast to meet 508 compliance. The names are also added as required for 508 compliance. We have changed the arrow to be an equal (=) sign.
TEP Peer Reviewer #1	Methods	The bias assessment is one of the features of systematic review that is a weak spot (in general). The authors address it by utilizing standard criteria and highlighting the bias in Tables of strength of evidence. Would the authors consider adding more methodology for the bias assessment? Specifically, how were disputes resolved? What is the scaling of the bias assessments (low, medium, and high) and how consistent were the bias reviewers?	We have added a few sentences describing the risk of bias assessment. We now state, "We resolved differences between reviewers through consensus. We judged the overall risk of bias for each study based on the adjudicated ratings for the individual risk of bias items. RCTs had three overall ratings for risk of bias (low, high, and unclear) and observational studies had five overall ratings (low, moderate, serious, critical, and no information)."
TEP Peer Reviewer #1	Overall	Do the authors find the term, 'sitter' offensive? While an accurate description of what many do, the aspirational goal is that these 1:1 personal attendants should be engaged in non-pharmacological delirium strategies. Would the authors consider modifying the term in the report?	Thank you for this thoughtful question. We agree that this term is unhelpful and we have changed it to "personal safety attendant" throughout the document.
TEP Peer Reviewer #1	Results	Would the authors consider commenting on the wide variability in the measurement of neurologic side effects? For example, Figure 9 identifies the EPS range from 2-10% - suggests some variability in measurement. Would the authors consider a statement about the variability? While the studies are at low bias for primary outcome, might there be some reporting bias for secondary outcomes?	Thank you to the reviewer for another thoughtful point. In response we have included the following points in the Discussion on page 146: "*Symptom reporting within the extrapyramidal designation was heterogeneous and likely represented a wide variety of measurement methods. Furthermore, studies included a variety of doses, frequencies and routes of antipsychotics, potentially obscuring the true frequencies of adverse neurologic side effects at the higher antipsychotic exposures."*
TEP Peer Reviewer #1	Discussion	In the discussion, the Contrast to the SCM guidelines gets nit picky, when it comes to English vs. non-English. The prior sentences suffice to draw the contrast. (page 146)	We have removed the sentence that the reviewer refers to in this comment.

Commentator & Affiliation	Section	Comment	Response
TEP Peer Reviewer #1	Overall	The report is clinically meaningful The target population and audience are explicitly defined The key questions are explicit and clear	Thank you for reviewing our report.
TEP Peer Reviewer #1	Introduction	The introduction is adequate	Thank you for reviewing our report.
TEP Peer Reviewer #1	Methods	The introduction is adequate	Thank you for reviewing our report.
TEP Peer Reviewer #1	Results	The results are clear - mostly	Thank you for reviewing our report.
TEP Peer Reviewer #1	Discussion/ Conclusion	The discussion needs one clarification “In the discussion, the Contrast to the SCM guidelines gets nit picky, when it comes to English vs. non-English. The prior sentences suffice to draw the contrast. (page 146)”	We have removed the sentence that the reviewer refers to in this comment.
TEP Peer Reviewer #1	Overall	Overall the report is incredibly well structured - please see comment about the 'Overall' section. The main points are clearly stated.	Thank you for reviewing our report.
Peer Reviewer #1	Overall	Rising awareness of the high prevalence and devastating consequences of delirium in the inpatient setting has highlighted the potential role of pharmacotherapies for delirium. Given this attention, this report on antipsychotics to prevent and treat delirium is timely and meaningful, and belongs in every inpatient providers’ literature library. The authors explain clearly the goal of the review, target populations, and key questions. The review is clear, well written, organized, and thorough.	Thank you for reviewing our report.
Peer Reviewer #1	Overall	Major Comment: While the review adds greatly to the literature, a potential downside it presents as being a “guideline” for antipsychotics is that busy providers, pharmacists, nurses, and administrators may quickly skim the Key Messages, Abstract, and Executive Summary and walk away thinking “antipsychotics are not recommended” or “antipsychotics are bad” rather than “there is no strong evidence supporting antipsychotics SPECIFICALLY for the prevention and treatment of delirium”, “in general, antipsychotics are safe,” and “more rigorous research is needed to define the role of antipsychotics for delirium.”	We agree with the reviewer that the message is a nuanced one. To that end we have attempted to communicate this by emphasizing what we have evidence for and being careful not overstate what we may not yet know. For example, as noted in our response to Comment #10, studies are heterogeneous enough in exposure to antipsychotic dosage to obscure the side effects, particularly at higher doses. While we were interested to note how few side effects were reported, we remain circumspect when emphasizing their safety based upon this evidence base.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Overall	In my experience, providers tend to have strong and conflicting opinions regarding antipsychotics, including efficacy, indications, mode of administration, dosing, and side effects. Some providers prescribe antipsychotics liberally to help patients stay calm, transition off continuous sedative medications, and sleep, while others quote antiquated "black box" warnings whenever antipsychotics are prescribed. Moreover, most hospitals are far behind in delirium practices, including management, lack of knowledge/expertise, inconsistencies in measurement, and minimization of deliriogenic practices (i.e., oversedation). Because of these factors, if misinterpreted, this review could lead providers, units, and/or hospitals to abandon antipsychotics altogether (i.e., not just for delirium).	We agree with most of the reviewer's comments. We do appreciate that misinterpretation (particularly to reinforce a pre-existing bias) can occur when users read a systematic review. We have carefully synthesized the available evidence and conveyed messages to minimize misinterpretation.
Peer Reviewer #1	Overall	Since the jury is still out regarding antipsychotics, and this review suggests that they indeed safe, it may be prudent to emphasize strongly in the Key Messages, Abstract and Executive Summary that this review SPECIFICALLY involves antipsychotics for delirium, and NOT for other potential uses such as sedation, agitation, and sleep. Moreover, greater "bottom line" emphasis should be placed on the safety of antipsychotics and the fact that this review is in no way discouraging use of antipsychotics, but merely summarizing the evidence. It would be unfortunate if misinterpretation of this review tightens the rules regarding antipsychotic prescribing, leading to a rise in use of more unsafe and potentially deliriogenic medications (i.e., benzodiazepines).	We are not as confident as the reviewer that these medications are "safe." Proving the absence of harms is a much more difficult endeavor and is prone to Type II error than demonstrating a difference when it is present. To stress that physicians need to be cautious in using these medications, despite the apparent low frequency of harms, we have now included a sentence which states that: *We did not detect serious neurological harms associated with haloperidol or second-generation antipsychotics used for the prevention or treatment of delirium, but cardiac effects tended to occur more frequently in antipsychotics compared with placebo.*
Peer Reviewer #1	Executive Summary	Abstract, introduction: As this review mostly found no effect of antipsychotics on delirium, the reader may infer that the included studies were heterogeneous or lacked methodological rigor. Consider addressing this in the abstract.	Thank you for reviewing our report. We have tried to make clear throughout where there is evidence for lack of effect versus a lack of evidence about an effect (such as due to limitations in quality or quantity of studies).
Peer Reviewer #1	Executive Summary	Abstract, introduction: The concluding sentence of the Abstract > Conclusions suggests that measures of patient agitation in the papers varied, but this is not clear in the Abstract > Results.	Thank you. We have deleted the words "agitation and", for currently the measures of patient distress seem to be lacking in the literature.
Peer Reviewer #1	Introduction	A theoretical/scientific/pharmacological basis (i.e., neurotransmitter pathways) for why antipsychotics are believed to prevent and/or treat delirium would add to the educational value of this section.	We agree that this is a topic of merit and interest but we are concerned that this would increase the length without greatly enhancing the messages of the report.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Introduction	Page ES-1 line 18: Delete the second “are”?	Second “are” is deleted – thank you.
Peer Reviewer #1	Introduction	Consider highlighting the heterogeneity in modes of evaluating delirium, and consider mentioning that standardized approaches now exist.	We agree with the reviewer that this discussion is meritorious but would add to the overall length without directly pertaining to the outcomes of the report.
Peer Reviewer #1	Methods	The methods are clear and logical, including inclusion and exclusion criteria, search strategy, outcome measures, and statistical methods. My only comment for the Methods is that the text in the Analytic Framework figure is a bit grainy and hard to read.	We have pasted new Analytic Frameworks into the report. Hopefully these are less grainy.
Peer Reviewer #1	Results	Evidence Summary > Results: Consider mentioning total sample size between studies.	In an earlier version of this report particularly when we were unable to calculate a summary measure, we detailed the sample sizes and description of individual studies. We were encouraged to increase the readability of the report by taking these details out of the text.
Peer Reviewer #1	Results	Evidence Summary > Adverse Effects: Consider opening this section with a short summary of adverse cardiac and neurological effects commonly believed to be associated with antipsychotics.	We have included this description in the Introduction of the Main Report. We do not think that a description of beliefs is warranted in the Results section of the Main Report of Evidence Summary.
Peer Reviewer #1	Results	Several studies also involve dexmedetomidine (Abdelgalel, Bakri, Carrasco, Konkayev, etc) as the only other therapy haloperidol is compared against. Perhaps consider changing the header of those sections (page 15 line 39, page 23 line 38, page 26 line 29, page 28 line 34, page 34 line 9, page 62 line 27, page 89 line 9, page 107 line 19) from “Haloperidol Versus Other Therapies” to “Haloperidol Versus Dexmedetomidine”. While this review is clearly not about dexmedetomidine, given its widespread use including in head-to-head trials, the authors may consider a brief mention of this medication in the executive summary.	Although an interesting suggestion changing the heading throughout the document would not be consistent with what evidence we sought or reflect the intent of the review.
Peer Reviewer #1	Results	Side effects – if not clear elsewhere in the results, readers may be curious to know the dose range of antipsychotics being given, to demonstrate that antipsychotics are usually safe even at higher doses.	Dose ranges for all of the studies are included in the Appendices in the Evidence Tables.
Peer Reviewer #1	Results	Page 15, line 45: dexmedetomidine is misspelled	Thank you. We have made this correction.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Discussion/ Conclusion	Discussion is well written. See Major Comment above and consider modifying Discussion to address that point.	Thank you for your comments.
Peer Reviewer #1	Clarity and Usability	See above comments.	Thank you for reviewing our report.
TEP Peer Reviewer #2	Overall	This is a well-written report that unfortunately highlights the lack of well done research into this topic. This review is unable to answer many of the pre-determined questions.	Thank you for reviewing our report.
TEP Peer Reviewer #2	Introduction	Well written	Thank you for reviewing our report.
TEP Peer Reviewer #2	Methods	Yes. The criteria selected for inclusion and exclusion seemed reasonable and based on a knowledgeable reader's experience with the literature. However, this project reveals little comparability between the studies and how much more work needs to be done in order to answer these questions.	Thank you for reviewing our report.
TEP Peer Reviewer #2	Results	Well written and described. Tables are somewhat redundant in places, where evidence is inadequate	Thank you for reviewing our report.
TEP Peer Reviewer #2	Discussion/ Conclusion	Major findings are clearly stated. More research is clearly necessary to fully understand prescribing of antipsychotics for persons with delirium.	Thank you for reviewing our report.
TEP Peer Reviewer #2	Clarity and Usability	Very well done. two typos noted P 45. Line 10 – mis-spell “effect” P 77 Line 29 - the sentence has no end	We couldn’t find the typo on page 45 but we did find “effect” misspelled on Page 19.
Peer Reviewer #2	Overall	I felt that the key questions were relevant and appropriately stated. These are common cohorts with unique aspects that need to be considered.	Thank you for reviewing our report.
Peer Reviewer #2	Introduction	Introduction: Well written	Thank you for reviewing our report.
Peer Reviewer #2	Methods	Methods: Again.. felt this was very comprehensive.	Thank you for reviewing our report.
Peer Reviewer #2	Results	Results: Appreciated the tables and forest plots.	Thank you for reviewing our report.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #2	Discussion/ Conclusion	I feel that the investigators did a good job with looking at relevant literature.	Thank you for reviewing our report.
Peer Reviewer #2	Clarity and Usability	Consistency between sections was helpful but at time due to length of report.. had to review multiple times due to limitations.	Thank you for taking the time to review this report. Systematically reviewing the different populations of patients has resulted in a very complex document.
TEP Peer Reviewer #3	Overall	This report is a valuable resource to clinicians and researchers. Most of my comments relate to readability and interpretation, with one minor comment on reproducibility (addition of objective detail in rating SOE). In general the authors executed the protocol appropriately and in accordance with feedback from expert panels. Results appear consistent with data presented.	Thank you for reviewing our report.
TEP Peer Reviewer #3	Overall	In the executive summary, background, 2nd paragraph: 1st sentence is confusing as written; perhaps replacing "are" with "...its..." makes more sense (pg ES-1, line 19-20).	The word “are” has been deleted.
TEP Peer Reviewer #3	Overall	In executive summary and throughout, when describing incidence reporting, text fluctuates between terms 'incidence' and 'occurrence'. Suggest use of incidence throughout.	We have replaced the word “occurrence” with “incidence” in the body of the Executive Summary. We have left the word “occurrence in the Key Messages in keeping with “plain language” standards.
TEP Peer Reviewer #3	Overall	From a messaging perspective, the discussion section should elevate the paragraph identifying the paucity of data available to make recommendations as a key finding of this review. Secondly, statements made about the potential for 2nd gen. antipsychotics to reduce delirium incidence in post-operative populations should be paired with a statement about the strength of the current evidence. As stated by the authors in the discussion, further research is suggested, and we must be cautious about readers who may change clinical practice prematurely.	Thank you for this comment. We have put a qualifier statement in the discussion regarding the need for further study regarding second-generation antipsychotics and the prevention of postoperative delirium.
TEP Peer Reviewer #3	Introduction	Figure 1 in this section does not include incidence as an intermediate outcome; would be prudent to add.	We have added delirium incidence to the analytic framework for KQ1.

Commentator & Affiliation	Section	Comment	Response
TEP Peer Reviewer #3	Methods	It seems odd that the report includes no observational studies describing the inappropriate continuation of antipsychotics after ICU/hospital stay. While a handful exist, I wonder if there was an exclusion criteria that removed these papers (such as the lack of validated delirium assessment)? Nonetheless, some response to this, or commentary on what has been reported in other studies may be worthwhile (i.e. 25-30% of antipsychotics used in the ICU/hospital are continued after discharge).	Our literature search identified the papers that the reviewer makes reference to. However, none of these studies met our inclusion criteria. The following reasons resulted in exclusion: the study being retrospective, no use of a valid delirium assessment tool, or antipsychotics not being used specifically for the treatment or prevention of delirium. Many of the studies evaluated the use of antipsychotics more broadly reporting on the inappropriate continuation of them in the general population of patients. Since these studies were not specifically about the prevention or treatment of delirium they were not included in this report.
TEP Peer Reviewer #3	Results	In general, results reported are valuable, and represent (in my opinion) an appropriate balance of summative detail from several studies conducted with great heterogeneity in population selection, intervention delivery, and outcome assessment.	Thank you for reviewing our report.
TEP Peer Reviewer #3	Results	One approach in the presentation of results stood out to me that may cause readers to pause, or could impact interpretation, is the variability in presenting RR's and SOE's inconsistently. It isn't clearly stated, but the SOE's are likely presented only for the outcomes deemed by the informants to be clinically relevant. However, one major point is that the only positive finding with respect to improvement from antipsychotics in delirium outcomes (i.e. reduction of delirium incidence in post-operative populations) does not include a report of the SOE. At the risk of readers changing clinical practice prematurely, reporting of this result should be accompanied by a statement of the SOE.	We pre-determined which outcomes would be graded by asking our Key Informants and Technical Experts which outcomes are the most relevant in making decisions about the use of antipsychotics in the prevention or treatment of delirium for each patient group. Using positive results to determine which outcomes to grade could potentially introduce some bias into our review. We reviewed the outcomes where we did not grade the body of evidence (did not do SOE) to ensure that we provide appropriate context about the strengths and limitations of the body of evidence.
TEP Peer Reviewer #3	Results	It is also relevant to consider adding more text in the discussion (particularly the executive summary) into the sub-populations studied (cardiac vs. orthopedic vs. other) to qualify this result and comment on future work needed to improve the quality of evidence.	We have added a phrase that qualifies the limitations of the evidence in the Evidence Summary and Abstract.
TEP Peer Reviewer #3	Results	The strength of evidence tables (6-24) include a column titled "Study Limitations" with low & medium indicators. It is unclear to me what low and medium mean, so perhaps a footnote describing response options and their definitions would be helpful, or revising the title of the column to be clearer to readers.	We have added a sentence to the Methods chapter describing study limitations. We state, “When scoring study limitations, we considered how well the study design and the study methods (using risk of bias assessment) protected against bias.”

Commentator & Affiliation	Section	Comment	Response
TEP Peer Reviewer #3	Discussion/ Conclusion	Discussion section highlights important challenges and qualifications of the findings. As noted in other comments, the second sentence of the conclusion should be tempered with a note on the strength of evidence/data related to 2nd generation antipsychotics and prevention. It isn't explicitly stated (but perhaps could be) that while the 2nd gen AP's may prevent delirium, they have not been shown effective as a treatment. Whether different mechanisms are at play is possible, but a conundrum nonetheless that should be clarified before clinicians react to existing data.	We have added this qualifier to the second sentence of the Conclusions as recommended. The text now reads "but this evidence is limited and requires more study."
TEP Peer Reviewer #3	Discussion/ Conclusion	The comment noted on page 147, line 39, is the first to identify existing data for inappropriate continuation of antipsychotics. It seems relevant to expand this comment, noting both why these observational studies were not included, and perhaps more detail into their findings so readers have some understanding of risks.	We have included a phrase that explains why the studies were not included in this systematic review.
TEP Peer Reviewer #3	Clarity and Usability	In general, the structure of the report is well organized, and follows the detail required to interpret this vast field of work. I feel that the authors should highlight earlier in the document (particularly the executive summary), that for many of the planned study questions, there was no evidence or low quality evidence to make recommendations. This in itself is an important finding that I believe is not transparent.	We have included a sentence in the Discussion of the ES at the start of the 3rd paragraph stating: *For the vast majority of outcomes predetermined to be of critical importance by our panel of experts and key informants, studies did not exist or were insufficient in design or number to answer the key questions.*
TEP Peer Reviewer #3	Clarity and Usability	Second, while the definition for 'Strength of Evidence' used in this report, and likely agreed upon by the authors in planning the report, it does have a degree of subjectivity in it, that may compromise reproducibility. Could the authors give some additional detail into any objective criteria used to delineate differences between low, moderate, and high strength of evidence (such as more than 3 high quality RCT needed for high SOE, and so on)?	We did not follow strict rules or counting of studies or dimensions to determine the strength of evidence. Grading a body of evidence is based on judgements and has a degree of subjectivity. We determined our certainty or confidence in the body of evidence by considering the domains outlined in the AHRQ Methods Guidance, which is very similar to the approach developed by others such as GRADE. The definitions of the levels of certainty in the evidence are provided in the Methods section. They are also now included as footers for the summary tables in the Evidence Summary.
TEP Peer Reviewer #3	Clarity and Usability	Third, while delirium severity was an agreed upon outcome of importance from the TEP and Expert informants, it seems odd that delirium duration is not reported in a similar manor.	The methods section outlines how we determined which outcomes would be graded. This was based upon voting by the TEP and Expert Panel. In response to this reviewer we have added text to provide context regarding the consistency and strength of each of the findings, such as duration.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer #1 (Medicus, Jennifer [APA])	Executive Summary	Figures A and B may need additional footnotes, explanation, and/or clarification in the text, particularly as the executive summary may be the only part that is read. For those that aren't familiar with AHRQ reports or GRADE methodology, the phrase "strength of evidence" is often erroneously interpreted as indicating the magnitude of a treatment effect, when in fact the analysis shows no difference between the intervention or control.	We added a footnote to Figures A and B of the Executive Summary defining each evidence grade.
Public Reviewer #1 (Medicus, Jennifer [APA])	Executive Summary	In wording the finding on treatment in palliative care settings and in displaying the finding in Figure B, the language is potentially confusing. The text states: For those being treated for delirium in palliative care or hospice settings, haloperidol or second-generation antipsychotics may have slightly less improvement in delirium severity than those treated with placebo (low strength of evidence)." However, it seemed to reviewers that the negative symbol in the table next to delirium severity would imply that delirium severity was reduced by the antipsychotic. The concept of "slightly less improvement in severity than with placebo" was hard to conceptualize without careful reading.	We have changed the sentence to read, "Patients being treated with haloperidol or second-generation antipsychotics compared with those who received placebo in palliative care or hospice settings may have slightly less improvement in delirium severity over time (low strength of evidence)."
Public Reviewer #1 (Medicus, Jennifer [APA])	Executive Summary	Not only do the tables have an inordinate focus (by wording and by block color) on the strength of evidence, but they entirely ignore the magnitude of the effect. In addition to being confusing (as noted above), the symbols related to conclusions only denote the direction of the effect and say nothing about its magnitude.	We have opted not to include the effect sizes in the summary tables because the effect size for some of the outcomes cannot be summarized simply. We reviewed the Evidence Summary to ensure that the effect sizes are included in the text, where available.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer #1 (Medicus, Jennifer [APA])	Executive Summary	Additionally, the presentation of the adverse effect information was confusing. The text states: "Apart from a single RCT in patients receiving palliative care which reported a statistically significant increase in extrapyramidal symptoms for both haloperidol compared with placebo and for second-generation antipsychotics compared with placebo, the larger body of evidence in all other patient populations, found no statistically significant increase in any neurological effect for any first- or second-generation antipsychotic compared with placebo or other head-to-head trials." This sounds as if neurological side effects including EPS were not problematic. However, it also states that "Neuroleptic malignant syndrome and various manifestations of extrapyramidal symptoms were the most commonly reported neurological effect." If neuroleptic malignant syndrome was being commonly observed in either group, that would be worrisome. We suggest being much more specific about what the studies actually showed.	Thank you for your careful review of the report. We have edited the statement by taking out specific mention of NMS. (see ES-6) To the same end, the main report: page 34 has now been edited to clarify the occurrence of NMS as follows: *"The two RCTs described above monitored for the occurrence of neuroleptic malignant syndrome. One RCT reported no statistically significant between-group difference in patients randomized to intervention and placebo for 3 mg and 6 mg daily intravenous dosage arms, respectively, finding that 2 cases of suspected neuroleptic malignant syndrome occurred in patients randomized to placebo.* The main report includes significant details of types of EPS symptoms and specifically mentions the occurrence (or lack thereof) of NMS on pages 34, 35, 51, 66, 67, 108-110.
Public Reviewer #1 (Medicus, Jennifer [APA])	Introduction	The scope, key questions, and analytic framework note that all first generation psychotics are included in the review, but there is no explanation in the introduction or methodology as to why the reported data only included haloperidol. If there were no available studies using other first generation antipsychotic agents, that should be reported in the text.	We added a statement, "We found one study that compared haloperidol with another first-generation antipsychotic. We decided not to discuss this comparison further because of the infrequent use of chlorpromazine in current clinical practice."
Public Reviewer #1 (Medicus, Jennifer [APA])	Methods	We recognize that input on the draft included public comments as well as input from peer reviewers and key informants, however, we note the Technical Expert Panel did not include a psychiatrist although they are frequently consulted for patients experiencing delirium and treat those patients for their symptoms. It may have been helpful to include the perspective of psychiatrists on the TEP to produce a more useful review.	The Principal Investigator of this report is a psychiatrist and a delirium expert.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer #1 (Medicus, Jennifer [APA])	Methods	In discussing the synthesis of the data, the draft report notes that "Since we anticipated that most drugs within a class would have similar physiologic effects, we combined studies of unique medications within classes when reporting outcomes." This approach seems problematic to us because the second generation antipsychotic medications have quite different pharmacological properties, side effect profiles and receptor-mediated actions. For example, as noted on the Neuroscience-based Nomenclature site, which is increasingly being used by journals to categorize psychotropic medications, olanzapine is a dopamine D2 and serotonin 5-HT2 receptor antagonist whereas aripiprazole is a D2 and 5-HT 1A partial agonist as well as a 5-HT2A antagonist.	We added an acknowledgement of the different mechanism of actions. We now state, "Although the drugs may have different mechanisms of action, we anticipated that most drugs within a class would have similar clinical effects. Therefore, we combined studies of unique medications within classes when reporting outcomes." As noted in the methods, we explored the effects of pooling unique medications if there was substantial heterogeneity.
Public Reviewer #1 (Medicus, Jennifer [APA])	Methods	Several reviewers noted concerns about the choice of primary outcomes to focus on for the analysis. The outcomes for the prevention and treatment studies are identical although it would seem that the focus should be different. For a prevention study, the incidence of delirium would be a crucial outcome and it would be helpful to have additional analysis and to distinguish between prevention of hyperactive delirium (e.g., with agitation) and hypoactive delirium, which is common but often unrecognized.	We pre-determined which outcomes would be graded by asking our Key Informants and Technical Experts which outcomes are the most relevant in making decisions about the use of antipsychotics in the prevention or treatment of delirium for each patient group. By following this process, we graded the outcomes that were deemed to be the most clinically-important and relevant to patient and caregivers. We have added text throughout Results section to provide more context about evidence for all outcomes.
Public Reviewer #1 (Medicus, Jennifer [APA])	Methods	We are also concerned that the other end points the review uses to measure success, such as length of hospitalization stay, are potentially misleading in terms of treatment impacts. When delirium is present, it often reflects the patient's underlying degree of illness and it is this underlying illness and not delirium per se that will often influence the length of hospital stay.	We agree with the reviewer that a patient's length of stay can be confounded by other variables such as illness severity and comorbidity. It is for this reason that RCT's are so valuable in determining the potential impact of an intervention compared to placebo.
Public Reviewer #1 (Medicus, Jennifer [APA])	Methods	From a measurement standpoint, the length of stay conclusions may be affected by floor effects. Lengths of stay are usually short anyway and any improvements would be more difficult to demonstrate. As noted in the draft report, the data on length of stay were skewed in their distributions and usually reported as a median making meta-analyses impossible. These factors may confound interpretation of the length of stay findings.	While we agree that the length of stay data are skewed and do not yield a summary measure, inferences can still be made from low risk of bias studies with large numbers of participants, particularly when they report the same findings. For example, such was the case for haloperidol vs. placebo in delirium prevention where we found 5 RCTs indicating no difference in the length of stay. (see Page 19, Full Report)

Commentator & Affiliation	Section	Comment	Response
Public Reviewer #1 (Medicus, Jennifer [APA])	Methods	Although some studies did look at agitation (which typically drives use of an antipsychotic in individuals with delirium), it would be useful to know whether the studies of delirium severity reported any agitation subscale information (e.g., with the Delirium Rating Scale-revised-98).	We collected data on agitation under short-term delirium symptoms. We added a footnote to Table 2 describing the symptoms considered short-term delirium symptoms.
Public Reviewer #1 (Medicus, Jennifer [APA])	Methods	We recognize that the body of evidence is assessed using the methodology delineated in the Guide for Conducting Comparative Effectiveness Reviews. Nevertheless, it is often unclear why strength of evidence has been rated as insufficient when studies may be available or as low when there may be one or two high quality RCTs related to an outcome. Additional transparency would be helpful for guideline developers who wish to use the information in the report to develop clinical recommendations.	We described the limitations to the body of evidence throughout the Results chapter, and footnote the reasons for downgrading the evidence in the strength of evidence tables.
Public Reviewer #1 (Medicus, Jennifer [APA])	Results	After reviewing the subpopulations results, it appears that much of the data comes from the over 65, post-surgery, or ICU populations (and in several studies combinations of these groups). When reviewing the subpopulation results, the text indicates that the entirety of the discussion in the overall analysis covers the subpopulation only. However, the overall discussion may lead a reader to assume that the results are applicable to any population with delirium; it may not be clear that the results in fact apply to a subpopulation only. We strongly recommend reorganizing the report so that outcomes based on subpopulations are reported in the subpopulation sections and the overall outcomes discussion sends a reader to the subpopulation discussion.	Although this is an interesting approach, our methodology was decided upon *a priori* and is presented as such. We are hopeful that the reader will read the full report, which details the types of studies that the overall information is derived from.
Public Reviewer #1 (Medicus, Jennifer [APA])	Results	Throughout the results, particularly for duration or length of stay outcomes, no meta-analysis was able to be run because of data being reported as median values. However, the text often states the AHRQ concluded that there was no effect of the intervention. Without a meta-analysis and studies that reported different effects, that seems like a strong and potentially misleading statement. If that conclusion is based on the majority of the studies reporting the same effect, that should be noted. An example is on page 19, for Duration of Delirium, Haloperidol vs Placebo, but also page 26 Length of Stay in Hospital Haloperidol vs Placebo. As one of the critical outcomes, these conclusions are of particular importance.	Thank you for this comment. We have edited the final sentence on page 19 as follows: '*Based on the consistent findings in six of seven trials we concluded that haloperidol has no effect on delirium duration when used as a preventive agent in all populations at risk of delirium."* Similarly we have edited the concluding sentence on page 26 as follows: *"Considering the consistent findings in six of eight trials, we concluded that there was no effect of haloperidol compared with placebo on the duration of delirium. (SOE: High)"*

Commentator & Affiliation	Section	Comment	Response
Public Reviewer #1 (Medicus, Jennifer [APA])	Results	Page 21, Figure 5. It's unclear why only three of the RCTs were included in the meta-analysis. It looks as if Wang 2012 was excluded as having an uncertain risk of bias, but in other meta-analyses (Figure 6), these types of studies were included. Also, it appears there is a typo and that Girard 2018 should be Schrijver 2018.	Thank you for your careful review – these errors have been corrected and the meta-analysis includes Schrijver, Girard 2010, Page and Wang. The combined point estimates and 95% confidence intervals have been changed in the text.
Public Reviewer #1 (Medicus, Jennifer [APA])	Results	Page 28, Length of Stay in ICU, Haloperidol vs Placebo – the last sentence should make clear whether none of the studies showed a clinically meaningful effect or whether the authors concluded there was no clinically meaningful effect based on a mix of findings.	We added: *Because all of the studies reported either no statistical difference, or no meaningful clinical difference we concluded that there is no effect on length of stay in the intensive care unit for all patients at risk of delirium when randomized to haloperidol compared* with placebo.
Public Reviewer #1 (Medicus, Jennifer [APA])	Results	Page 29, sedation haloperidol v placebo – in this paragraph as well as in multiple other places in the document, study information is not included in the meta-analysis because the number of events was low or zero. Even if such an approach is necessary on mathematical grounds, it seems misleading in terms of clinical conclusions. The absence of sedation with a given treatment is, in fact, important to know when assessing the potential benefits and harms of a treatment for guideline development purposes.	Thank you for this comment. We agree that it is helpful to know this and we have included a sentence on Page 29 to describe this study. It is precisely because of the rare events in relation to sedation that we have been unable to come to a firm conclusion based upon this evidence. We have revised the figure to incorporate the study that was excluded from the analysis due to no events.
Public Reviewer #1 (Medicus, Jennifer [APA])	Discussion/ Conclusion	Limitations of the Evidence Base – Based on the last two paragraphs starting with "Many studies were underpowered…", we believe it would be more helpful to acknowledge these shortcomings upfront and in the executive summary, and to soften the rather negative message about the utility of antipsychotics in treating symptoms, such as agitation, in patients with delirium. Sadly, the only really legitimate conclusion is that despite a lot of effort and money spent on these issues, the quality of the evidence does not allow one to rule in or rule out utility of antipsychotics for delirious patients.	We agree with the reviewer that the literature was insufficient to answer many of the questions regarding critical outcomes established a priori. We have stated this in the results section of the Executive Summary. We believe that the message of this report follows the findings from the data synthesis, conducted in accordance with AHRQ guidelines and as outlined in our protocol. Further research to evaluate use of these medications for very specific symptoms, such as agitation, may be valuable to clinicians; however, making this type of statement, within the report, is outside of its intended scope.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer #1 (Medicus, Jennifer [APA])	Discussion/ Conclusion	Research Recommendations – We recognize that the scope of this document was limited to the effects of antipsychotic medications in individuals who have or are at risk for delirium. Among this group of patients, however, it would be important to determine which specific aspects of delirium (if any) are affected by antipsychotic treatment. It would also be important to determine whether patients with hypoactive delirium are being miscategorized as having daytime somnolence and whether antipsychotic medications have different effects on hyperactive as compared to hypoactive delirium.	We agree with the reviewer that more study is needed of patients with delirium and the role of antipsychotics in reducing distress and agitation. We have included the following sentence in the ES-7 conclusion sentence as follows: Future studies should include standardized, clinically meaningful measures of patient agitation and distress, subsequent memories of delirium, caregiver burden and distress, inappropriate continuation of antipsychotic therapy, and long-term cognitive and functional outcomes.
Public Reviewer #1 (Medicus, Jennifer [APA])	Discussion/ Conclusion	Studies are also needed of preventive interventions in non-surgical patients.	We agree. We have included all that we could find in the literature in this review.
Public Reviewer #1 (Medicus, Jennifer [APA])	Discussion/ Conclusion	Research publications and systematic reviews should also do a better job of reporting the representativeness of the patients being studied. In many clinical trials, the most seriously ill individuals are not included, which can bias findings. Individuals with severe agitation in the context of hyperactive delirium and those with serious medical instability (which is also common with delirium) may not be included in the sample. This reduces the relevance of any study findings. At the very least, reporting such data is crucial for readers to understand the patient population that is being studied. Known risk factors for delirium such as pre-existing cognitive impairment should also be reported. Among post-operative patients, the type of surgery and information on surgical parameters (e.g., operating room times, physiological changes during surgery) should be described more clearly.	Thank you for this insight. We agree that more standardization in the reporting of results in trials will allow better understanding of generalizability of findings. For this reason we have included in our discussion the need for increased standardization in delirium studies.
Public Reviewer #1 (Medicus, Jennifer [APA])	Discussion/ Conclusion	It is also crucial for research to focus on aspects of delirium unrelated to antipsychotic treatment, per se. Unfortunately, much of the research has misguidedly focused on trying to prevent and treat the manifestations of delirium with medications. Instead, greater attention should be given to proactive identification and treatment of factors that contribute to the development of delirium (e.g., physiological changes or abnormalities, medication related issues) as well as to enhanced screening approaches to identify delirium. In prior studies, almost half of patients with delirium are never identified and this is a pre-requisite to any further intervention.	We agree with the reviewer's comments that screening, appropriate work-up to determine the underlying cause of a delirium and non-pharmacologic approaches to delirium, are all critical to best practices in delirium care and prevention.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer #1 (Medicus, Jennifer [APA])	Discussion/ Conclusion	In studies of delirium identification and treatment, it is also important to incorporate frequent longitudinal measures. By its very nature, the observed symptoms of delirium fluctuate as does the patient's level of consciousness and infrequent cross-sectional measurements can be highly misleading. Sleep wake cycle disruption is another common sign of delirium and agitation may be more prominent at night when research assistants are less likely to be performing rating scales.	We agree that fluctuation can make delirium symptom measurement and aspects of sleep wake cycle disturbance difficult to measure.
Public Reviewer #1 (Medicus, Jennifer [APA])	Overall	The APA will begin work on a clinical practice guideline on delirium at the end of 2019. Ultimately, our goal for any systematic review is to be able to use it as a basis for writing our guidelines. Unfortunately, the overall scope and aim of this review will make it impossible to use in writing a practice guideline.	Thank you for taking the time to review this report.
Public Reviewer #1 (Medicus, Jennifer [APA])	Overall	In our experts' clinical experience, although antipsychotics may or may not prevent delirium in some patients, they definitely do not "treat" delirium. Instead, antipsychotics are used to manage the hyperactive behavioral symptoms of delirium, particularly extreme anxiety, suspiciousness, psychosis, agitation, and aggression (e.g., patients pulling out lines and monitors, hitting care providers, thrashing about). Treatment needs to correct the underlying causes of delirium, whether eliminating deliriogenic medications, treating infections, correcting metabolic imbalance, or other etiologies. At the very least, we would suggest changing the title of the review to "AHRQ Draft Report on Antipsychotics for the Management of the Symptoms of Delirium" to help clinicians appreciate this important distinction. It would also be valuable to emphasize these points in the background that is provided at the beginning of the document.	Thank you for this insightful comment. We believe that the findings of this review begin to underscore this viewpoint. However we do not agree with changing the title of the report, as the current title accurately reflects the primary objectives of the systematic review.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer #1 (Medicus, Jennifer [APA])	Overall	It is concerning that the report describes delirium as a phenomenon "generally associated with underlying medical causes" (pp ES-1, 1, first paragraph for both), which makes it sound as if delirium were sometimes a primary psychiatric illness. When a workup does not reveal an obvious and dramatic medical cause of the delirium, the assumption of the medical/surgical team is often that there is no medical cause and they erroneously attribute it to a psychiatric condition. However, it is possible to have multiple abnormalities that synergistically result in delirium (e.g., a patient who simultaneously has a low hematocrit, reduced oxygenation, fluid/electrolyte imbalance, and infection). A careful overview from AHRQ of the wide range of delirium etiologies to look for and correct while managing the patient would be more helpful to clinicians. In addition to deleting the word "generally" in the phrase noted above, it should also be pointed out that it is essential to recognize delirium swiftly and to correct the precipitating cause, which can be the medical treatments themselves. Antipsychotics are merely a tool to manage the patient's symptoms, agitation, and distress while the underlying cause is identified and corrected.	We agree with all of these thoughtful comments. We have removed the word "generally" from the Background Sections in the Evidence Summary and in the Main report.
Public Reviewer #1 (Medicus, Jennifer [APA])	Overall	Many, if not most, of delirious patients actually have hypoactive delirium, and antipsychotics do not help these patients. It is highly likely this issue confounded results of the recent study published in NEJM by Girard et al. and it is also likely to be present in other study samples that have been used to test the efficacy of antipsychotics to prevent and treat delirium. However, in clinical care, psychiatric consults are usually only called for patients with hypoactive delirium if they are not cooperating with treatment. (In that case, the call is for a consult to assess capacity.) It is the patients who are agitated that generate the most concern among patients' families and hospital staff and in whom it would be helpful to know the best pharmacologic approach to take.	We agree with this comment and have included to following sentence: *Future studies should include standardized,* ***clinically meaningful measures of patient agitation and distress,*** *subsequent memories of delirium, caregiver burden and distress, inappropriate continuation of antipsychotic therapy, and long-term cognitive and functional outcomes."* as the last sentence of the Evidence Summary Conclusions.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer #1 (Medicus, Jennifer [APA])	Overall	One of the few positive findings of this report was that second-generation antipsychotics decreased the occurrence of delirium compared with placebo in patients at risk of delirium (RR, 0.36; 95% CI, 0.26 to 0.50); however, this meta-analysis (of 3 studies) only included postoperative patients. Although the draft notes that haloperidol had no effect on delirium occurrence across all studies, in this context it would be important to know whether haloperidol had an effect with an analysis that was limited to the post-operative studies.	Thank you for this question. As documented in the Main Report under Postoperative Intermediate Outcomes we have documented the four trials randomizing patients undergoing surgical procedures to haloperidol and placebo and have found a combined relative risk of 0.86 (95% CI, 0.576 to 1.29) suggesting that there is no evidence that haloperidol prevents delirium in a heterogeneous group of surgical patients.
Public Reviewer #1 (Medicus, Jennifer [APA])	Overall	There also needs to be more specificity throughout the document on the types of post-operative patients that were studied. One part of the document notes that post-operative patients had had hip surgery, GI surgery, and cardiac surgery, but each of those may be quite different from each other and from other surgical subgroups. Even in terms of GI surgery, elective bariatric surgery is likely to be quite different from GI surgery for trauma, cancer, or an ischemic or perforated bowel. Greater specificity of the study population in the summary sections of the document will help readers draw appropriate conclusions.	Due to space restrictions, we have to focus on the broad themes of the overall body of evidence and are unable to report on all the details of the studies in the main report. These details of the individual studies, including specifics of the patient populations, are included in the Evidence Tables in the Appendix. We have added notes referencing the Appendix following the Search Results.
Public Reviewer #1 (Medicus, Jennifer [APA])	Overall	Finally, there are multiple typographical and grammatical errors throughout the draft, which we have not noted specifically assuming there would be a final proofread of the text.	We have reviewed the report for grammatical errors.
Public Reviewer #2 (APA)	Overall	We appreciate the call for further research on the efficacy and effectiveness of antipsychotics for the prevention and treatment of delirium in older patients and critically ill populations.	Thank you for reviewing our report.
Public Reviewer #2 (APA)	Overall	The attention given to examining potential harms of medications is commendable given that the side effects of antipsychotics such as Haloperidol can be deleterious to the patient and his/her caregiver.	Thank you for reviewing our report.
Public Reviewer #2 (APA)	Overall	It is important to examine nonpharmacological interventions for the prevention and treatment of delirium, including psychological treatments.	We couldn't agree more with the reviewer.
Public Reviewer #2 (APA)	Overall	We appreciate the attention given to context, specifically the patient's environment in the key questions.	Thank you for reviewing our report.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer #2 (APA)	Overall	We appreciate the attention given to as well as the call for additional research into patient-centered outcomes such as quality of life, patient distress, and caregiver burden/distress.	Thank you for reviewing our report.
Public Reviewer #2 (APA)	Executive Summary	P. ES-1: Insert the bolded word in the sentence below located at the beginning of the second paragraph: Preventive and therapeutic interventions are needed to reduce the burden of delirium and are associated [with] long-term cognitive impairments.	Thank you – This sentence is now corrected.
Public Reviewer #2 (APA)	Executive Summary	P. ES-3: Antipsychotics for the Treatment of Depression: Delete the following word in the last sentence of the first paragraph. Critical outcomes by patient group ~~and~~ are listed in Figure B.	Thank you- this word is now deleted
Public Reviewer #2 (APA)	Executive Summary	Suggest further defining the first sentence of the discussion in the executive summary ("Though frequently used in patients with delirium or those at risk of delirium, there remains a lack of clear evidence to support their use") to indicate which medication is being referred to in this sentence.	Thank you for your careful review. This sentence now reads: "*Though antipsychotic medications are frequently used in patients with delirium or those at risk of delirium, there remains a lack of clear evidence to support their use for overall prevention or treatment.*"
Public Reviewer #2 (APA)	Results	P. 19 (Delirium-free, Coma-free Days Alive: Haloperidol Versus Placebo): There is a typo in the following sentence: We concluded that haloperidol had no effet on delirium-free or coma-free days alive.	Thank you- this error has been corrected.
Public Reviewer #2 (APA)	Overall	Suggest further explaining the phrase, "We did not detect…" when used to reference harms. Does this mean that there was insufficient evidence available about potential harms, or does this mean that information was available but indicated that there were no harms?	Thank you for this question. We have reviewed the entire report with this question in mind and have attempted to clarify this comment with each outcome.
Public Reviewer #3 (UNRO Organization, Universal Necessity Relief Organization)	Not Applicable	Received comments unrelated to the topic and the report.	Not applicable