Comparative Effectiveness Review Disposition of Comments Report

Research Review Title: First- and Second-Generation Antipsychotics in Children and Young Adults: Systematic Review Update

Draft review available for public comment from June 10, 2016, to July 8, 2016

Research Review Citation: Pillay J, Boylan K, Carrey N, Newton A, Vandermeer B, Nuspl M, MacGregor T, Ahmed Jafri SH, Featherstone R, Hartling L. First- and Second-Generation Antipsychotics in Children and Young Adults: Systematic Review Update. Comparative Effectiveness Review No. 184. (Prepared by the University of Alberta Evidence-based Practice Center under Contract No. 290-2015-00001-I.) AHRQ Publication No. 17-EHC001-EF. Rockville, MD: Agency for Healthcare Research and Quality; March 2017. www.effectivehealthcare.ahrq.gov/reports/final.cfm. doi: https://doi.org/10.23970/AHRQEPCCER184.

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 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.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 3	General	The report is important and clinically meaningful. The target populations are well defined, but the reason for not including Intellectual Disability as a condition of interest in inexplicable. There is more data on disruptive behavior in ID than for a number of the conditions included.	Thank you for this comment. The inclusion/exclusion criteria was considered by several Key Informants and Technical Experts in the original review and for this update, as well as posted for public review, during which time intellectual disability was not suggested as a discrete condition of interest. ID was not an exclusion criteria for studies in ASD (or other conditions), and was a key participant variable in some studies, and data and comments to this effect are included in the study description tables and in our sensitivity analysis for autism spectrum disorders.
Peer Reviewer 4	General	It is clinically meaningful and the target population and audience are identifies. The key questions are appropriate and explicitly stated.	Thank you for your comment. No response required.
Peer Reviewer 9	General	This is a well conducted systematic review to update the previous report on the effectiveness and harms of First- and Second-Generation Antipsychotics in Children and Young Adults	Thank you for your comment. No response required.
TEP 1	General	Yes, the report is clinically relevant. Target population, audience, and key questions are specified and appropriate.	Thank you for your comment. No response required.

Commentator & Affiliation	Section	Comment	Response
TEP 2	General	The results are clinically meaningful although somewhat confusing as the evaluation of SGAs individually and generally at times pointed to moderate effect on symptom reduction but little to no difference in CGI-I or CGI-S. Discussion of this difference by the authors would be of value.	This is a good point. We have tried to clarify the findings (when low or higher SOE) for the relevant conditions and outcomes (including response rates) with the following text: 1. Bipolar (ES, key points): (for quetiapine vs. placebo): "the results of little or no difference for response rates (often focused on manic symptoms) were imprecise showing that many patients may have clinically relevant response." 2. ASD (ES, Key points) (for individual SGAs): "smaller sample sizes contributing to the SOE for each drug likely affected the ability to obtain a significant finding for most outcomes (e.g., response rates), with the exception of irritability which overall had the larger magnitude of effect." 3. ADHD: (ES, Key Points) (all SGAs vs placebo) "Results for clinical impressions of improvement showed little or no difference, although results were imprecise and indicated that many patients may possibly improve."
TEP 2	General	The target population and the audience were well defined and the key questions were appropriate and explicitly stated.	Thank you for your comment. No response required.
TEP 5	General	Very thorough and useful report. The executive summary is the right length to highlight the salient findings. The target population and audience are clear as were the key questions. The outcomes and description of harm types were also clear and sensible. It was good to mention the lack of evidence for use in other mental health conditions and the lack of evidence around patient centered outcomes. Mentioning the lack of clinically meaningful difference and what degrees of change suggested them beyond statistical significance was a good reminder.	Thank you for these comments. No response required.

Commentator & Affiliation	Section	Comment	Response
TEP 5	General	The appendices were well organized and useful. In particular, the reason for excluding studies was helpful.	Thank you for your comment. No response required.
TEP 5	General	[page vii] Add first “generation antipsychotics”. Throughout the document, the authors use an “s” on SGA to represent second generation antipsychotics, so to be consistent, it should read SGAs.	Thanks. We have specified SGAs and FGAs in the abstract introduction.
TEP 5	General	[page vii] add “with the exception of aripiprazole and risperidone” SGAs “as a class”, [page vii] SGAs as a class (and risperidone individual); (and risperidone individually) would be listed, as it belongs to the SGAs’ class. [page vii] just added a space between words “less” and “than”	Thanks for the suggestions. We have incorporated most of these.
TEP 5	General	I see the references, abbreviations and acronyms section, but may want to consider defining all acronyms in the Structured Abstract ADHA, BMI, EPS the first time you use them in the structured abstract to be consistent with how FGAs and SGAs are treated.	Thanks, we have added the full terms.
TEP 6	General	An excellent report combining clear, concise explanations, solid methodology, and a reasonable presentation of the data. The key questions, population descriptions, and results and are explicit and clinically meaningful.	Thank you for your comment. No response required.
TEP 7	General	As a biostatistician, I am evaluating the report mainly on the basis of the meta-analytic methods and their interpretation. From this standpoint, I found the report to be excellent. Key questions were clearly stated, and the statistical methods were tailored to answer them directly. The most notable shortcoming was the lack of a detailed description of the meta-analytic models. It was difficult for me to evaluate the models without seeing their structure. The written descriptions of the models were not sufficient. I would recommend adding a statistical methods appendix where models can be specified, and if possible, the WinBUGS code used to fit the models. These are key components for making this research reproducible.	Thanks for this review and comment. The model structures have been specified in an appendix with WinBUGS code.

Commentator & Affiliation	Section	Comment	Response
TEP 10	General	Yes to all	Thank you for your comment. No response required.
Peer Reviewer 3	Introduction	The key questions are clear, but one important question that is not included is how the "harms" from atypical antipsychotics relate to each other. In particular how does weight gain or bmi relate to metabolic syndrome and risk for diabetes. If these factors correlate highly with weight gain or bmi then some of the lab monitoring might be most indicated in specific patients. This is a huge issue because drawing blood in some children/adults with disabilities may require sedation. It is critical that this issue be addressed.	We understand that correlation between "intermediate" and "final/patient important" outcomes is critical for influencing decisions including monitoring. This review did not specifically address these questions which would be excellent contributions to the evidence base. Apart from weight and BMI, we did extract data, when available, on a wide range of metabolic outcomes in an attempt to inform decisions on these "final" outcomes. Unfortunately there was limited data on many outcomes, but we were able to make some conclusions for some (incidence of diabetes, increased cholesterol and triglycerides) which should be informative for clinicians and other decision makers.
Peer Reviewer 4	Introduction	Background Information was well done.	Thanks for this comment. No response required.
Peer Reviewer 9	Introduction	The background is concise. Discussion of the different medical conditions is helpful. For key questions, the effectiveness were evaluated by each condition, and the harms were assessed across all conditions. The harms may not be impacted by the underlying condition, or too sparse to evaluate by condition, but it is helpful to state clearly why looking at harms across all conditions.	Thank you for this comment. We added a statement (ES and report Inclusion/Exclusion criteria sections) about the rationale for looking at harms across conditions, "The primary focus in KQ2 was harms across all conditions because adverse events associated with an antipsychotic are likely to be consistent regardless of the indication for which a drug is being taken; the difference in harms between conditions was treated as a subgroup of interest."
TEP 1	Introduction	Good overview.	Thank you. No response required.
TEP 1	Introduction	On page 16, line 10, there is reference to increased use of antipsychotics in children through 2002. Are more recent pharmacoepidemiology data available?	Thanks for pointing out this dated citation; we modified our text and reference (ES and text introduction; first paragragh) with prescription rates for 2010.
TEP 2	Introduction	No comments	No response required.

Commentator & Affiliation	Section	Comment	Response
TEP 5	Introduction	The introduction is clear and gives a good description of the problem and why it is important. The key questions are clear as is the analytic framework.	Thank you. No response required.
TEP 5	Introduction	Would it be worth putting in a few sentences as to how long FGA and SGAs are thought to reach a steady state effect? Because the length of studies 3 weeks to months is so different, it may help the reader put some context around the trial lengths. Similarly, how long does it take for adverse effects to develop, for example if tardive dyskinesia takes years to develop then one is less certain about the safety of long term use.	We agree with adding some additional explanation about the expected effects based on treatment duration. Steady state data may not be the most suitable because of the wide variety of dosing regimes (including fixed and flexible, lack of reporting of titration period) used in the primary research. We added a few sentences (Applicability section) we feel will help with interpretation though. "Adequate trials of antipsychotic treatment to assess response can be considered within 4 to 6 weeks, which supports applicability from the evaluated studies for these outcomes at least in the short term; nevertheless, issues impacting longterm treatment success, such as treatment compliance and resistance, were not accounted for in many studies… few studies allowed for conclusions on major adverse effects—especially those often arising with longterm treatment (e.g., tardive dyskinesias, diabetes). Adverse effects may have been underestimated due to the short followup periods; not all effects are likely to become evident in all patients within the one-to-two month treatment phase commonly conducted."
TEP 6	Introduction	Appropriately concise, explaining the context of the report.	Thank you. No response required.
TEP 7	Introduction	[page 23]: You say "pooled", but also that you used a random effects model, so this implies partial pooling. If effects are allowed to vary by study, then there is not complete pooling.	We have changed the term pooling with combining throughout to avoid confusion.
TEP 7	Introduction	[page 23]: "combing" should be "combining"?	This has been corrected.
TEP 7	Introduction	[page 23]: So, it sounds like you mean by "pooling" simply that you included them in the same analysis, and not "pooling" in the statistical sense? Please clarify.	We have changed the term pooling with combining.

Commentator & Affiliation	Section	Comment	Response
TEP 7	Introduction	[page 23]: "...evaluate a suite of comparisons using indirect evidence..."	We have added this clarification.
TEP 7	Introduction	[page 23]: Why not use study-level (or arm-level) covariates instead of stratifying?	We have clarified our approach to subgroup analyses, and mentioned that we conducted sensitivity analyses when having few studies in most cases (of benefit outcomes). We used study-level covariates in the metaregressions for harms because of the larger number of studies.
TEP 10	Introduction	Useful background. Probably helpful for the reader to know this is an update of a previous report...glad that is clearly stated.	Thank you. No response required.
Peer Reviewer 3	Methods	The inclusion and exclusion criteria are appropriate except I am not sure why papers on disruptive behavior in individuals with intellectual disability were not included.	Thank you for this comment. ID was not an exclusion criteria for the studies, and was a key participant variable in some studies ASD), and data and comments to this effect are included in the study description tables and in our sensitivity analysis for autism spectrum disorders.
Peer Reviewer 3	Methods	I think greater justification is need for the statement (ES-4) that confounding by indication is not an important threat in studies examining unanticipated harms. Many of the harms of antipsychotics are well know and influence prescribing practices.	We had been specific to how confounding is limited for many unanticipated harms in treatment naïve patients. We are not convinced that patient factors (apart from previous treatment harms) related to the differential harm incidence are well understood, and thus systematically bias results. The within-group subgroup analyses we've documented found little consistency in any effects.
Peer Reviewer 4	Methods	The search strategies and definitions as well as statistical methods used are appropriate or justifiable.	Thank you for this comment. No response required.
Peer Reviewer 9	Methods	Standard systematic review methods were used and the outcomes were clearly delineated. Calculation of standard deviation for continuous data, when not directly reported, adequately used different sources of data.	Thank you this comment. No response required.
Peer Reviewer 9	Methods	Specify the choice of change score vs. follow up score for continuous outcomes.	Good point. We have clarified when we used followup (SMD analyses) scores or (preferably) change scores.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 9	Methods	The review conducted both pairwise meta-analyses for outcomes with at least three studies and network meta-analyses for two outcomes. All analyses were conducted in a Bayesian framework.	No response required.
Peer Reviewer 9	Methods	For the analyses, the terms “aggregate” and “individual” may be confused with the terms of MA using aggregated study level data and MA using patient level data. Something like “across generation synthesis” and “individual drug synthesis” may be clearer?	We have reviewed our description of “aggregate (across class)” vs “individual FGA/SGA” “drug comparisons” and don’t feel that our readership will be confused especially since patient level data is not described or suggested. Indeed, no TEP or other external reviewer commented on any confusion.
Peer Reviewer 9	Methods	It is fine to use Bayesian method to conduct pairwise meta-analyses. The estimates from three studies are more stable than those from two studies, though the number of studies is still small for estimating variance parameters. In general the number of studies included in the metaanalyses were small, in particular, for the within drug class comparison.	We can understand the concern with variance parameters, and have added specifics in Appendix G about how priors were obtained for between studies parameters.
Peer Reviewer 9	Methods	For combining RCTs and NRCTs of effectiveness outcomes – do the NRCTs have similar baseline characteristics between groups? Did the studies provide adjusted estimates for effect size? Not sure about the fundamental differences between NRCTs and cohort studies in this context with medication interventions and they may share a lot of similar limitations here, but cross-design synthesis needs careful considerations and should not be just simply combined. If there are adjusted estimates, they should be used.	We agree that this could benefit from clarification. We have added a definition of what we considered a nonrandomized trial; only difference may be lack of randomization, yet allocation should be objective (i.e., not due to patient or provider preferences), and other biases could potentially be prevented, as with RCTs, to a greater extent than with observational studies. We have also mentioned our approach to combine studies and then explore a priori variables that may contribute to heterogeneity, as was done for all benefit outcomes (via sensitivity analysis) and via the regressions for the harms. We also documented all within study subgroup analyses to inform what may contribute to differences between studies should their study characteristics vary. Very few observational studies did any sort of adjustment especially for harms outcomes for which this design was used most.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 9	Methods	Agree less bias for combining harm studies. Still the comparability of study groups needs to be evaluated and considered.	We agree. Major differences arising from imbalances in baseline characteristics will often translate in heterogeneity between studies, especially between trials and observational. This is accounted for within the analysis (of between study variance), and assessment of the strength of evidence. There are no consistent differences between the results from trials and observational studies when visually inspecting forest plots; we carefully reviewed Figures 82-85.
Peer Reviewer 9	Methods	Methods for publication bias (small study effects) were mentioned but no results were presented in the results section?	We had placed these findings in the section on limitations of this CER, and to avoid adding too much length/repetition have left this here rather than moving to the results section.
Peer Reviewer 9	Methods	How zero events are handled in the Bayesian MA?	This information has been mentioned in the appendix material describing the analytical methods.
Peer Reviewer 9	Methods	Provide more description of the models and the considerations for the validity of the network meta-analysis.	We have added this information to the appendix material, and have extended the description in the report text (Methods), “We conducted convergence diagnostics (i.e., convergence verified using autocorrelation, paying particular attention to prior distributions on between study variance parameter) and assessed the fit of the models by monitoring the deviance parameters; the analyses were also checked for consistency by contrasting direct and indirect estimates in every closed loop of the networks with a display of the results in plots.”
Peer Reviewer 9	Methods	The models were run for 220, 000 iterations, instead of 2200, 000 iterations?	Yes they were run for 220,000 iterations; we corrected our typo. Thanks for pointing out.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 9	Methods	Specify the methods for model convergence diagnostic and assessment of goodness of fit.	We have added this material to the appendix and have extended the description in the report text (Methods), "We conducted convergence diagnostics (i.e., convergence verified using autocorrelation, paying particular attention to prior distributions on between study variance parameter) and assessed the fit of the models by monitoring the deviance parameters; the analyses were also checked for consistency by contrasting direct and indirect estimates in every closed loop of the networks with a display of the results in plots."
Peer Reviewer 9	Methods	Provide a reference for inconsistency factor plotting – The method is not called "inconsistency factor plotting", but it is "loop-specific approach" that evaluates inconsistency separately in every closed loop of a network of interventions, and the results are displayed in a plot. This method provides very low power to detect inconsistency. If the lower confidence interval limit of the inconsistency factor for a loop does not reach the zero line, it could be considered to present statistically significant inconsistency. However, the absence of statistically significant inconsistency is not evidence of consistency because of the multiple and correlated tests that are undertaken and the low power. The comparability of the studies in terms of potential effect modifiers should always be considered. It helps to employ more than one method. See more from Veroniki et al. 2014, BMC Medical Research Methodology 2014, 14:106.	Thanks for this clarity. We have modified our description of the approach, added the applicable reference as suggested, and ran an adjusted NMA using the variable of treatment duration to add assessment of inconsistency based on treatment duration - which was the only variable shown as a significant effect modifier in the other subgroup analyses on harms. We have also added some description of the limitations of the NMA in the discussion section (Limitations of this CER). "The findings from our network meta-analyses should also be considered exploratory in nature. Apart from the assumptions made for all meta-analyses, the network approach assumes transitivity, where we assume that all treatment nodes not present in any trial are missing at random, and there is nothing systematically different about the populations in the various trials. Because of these limitations we did not use these results for making our assessments of the strength of the body of evidence. We note, however, that the results were very similar to those found in the standard pairwise meta-analyses and the adjusted analysis factoring in treatment duration (shown as significant treatment modifier) did not change the results."

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 9	Methods	Not clear what it meant by "we also performed our own subgroup analyses using study-level data…" ? Clarify. Just "comparison" across studies?	Yes this is comparison across studies. We have provide clarity here, "we also performed our own "across study" subgroup analyses using study-level data on our variables of interest (e.g., phase of treatment, mean age of participants), where possible". Further sentences describe the methods to perform these, whether they were sensitivity analysis or regressions etc.
Peer Reviewer 9	Methods	Network meta-analysis is largely observational, too.	We agree, and added (ES and main report) "Findings from the network meta-analysis are considered fairly observational in nature and were compared with other more direct findings from pair-wise meta-analyses." Our additions to the Limitations section also expand on this.
TEP 1	Methods	The methods are sounds and standard for this type of analyses	Thank you for this comment. No response required.
TEP 2	Methods	Yes to each of the questions.	Thank you for this comment. No response required.
TEP 5	Methods	The inclusion and exclusion criteria are clear and make sense from a clinical perspective. The search strategy is clear, comprehensive and not likely to have missed much important information. Is it worth a comment as to the impact, if any, of not including non-English language studies? The outcome measures and the scales used to obtain them are clearly described.	Thank for this comment and considerations. Language has shown to not effect results to any meaningful degree & is considered more of a selection/language bias for types of interventions most often commonly used & studied in foreign countries (e.g. complimentary medicine), or when prevalence of the condition is much higher in these countries. We added a note to this effect, with 2 citations, in the section of limitations of the full report, "Moreover, effect sizes in language restricted reviews have shown to not differ significantly (overestimating effect sizes by 2 percent) from those not having restrictions.[212] Non-English publications are thought most important to seek for reviews of certain interventions, such as complimentary or alternative medicine, or when the prevalence of the condition or use of the intervention is particularly high in foreign countries.[212-213]"

Commentator & Affiliation	Section	Comment	Response
TEP 5	Methods	The statistical methods, though I am not a statistician appear appropriate. Is it worth mentioning why a Bayesian as opposed to a frequentist approach? What advantage does meta-regression using the Bayesian approach confer? I suspect many of the readers won't be as familiar with this approach.	We appreciate the concern. We have added a sentence justifying our use of an alternative approach to the traditional DerSimonian Laird (which may also be questionable to readers). Several alternatives to the DL approach are acceptable; we have use a frequentist approach (Hartung-Knapp-Sidik-Jonkman random effects model) in the past although found it very limited (highly imprecise) for meta-analysis with few studies.
TEP 6	Methods	Methods appear to be solid, logical, and explicitly stated. Similarly, definitions, statistics, and abbreviations are appropriate and are easy to follow.	Glad you think so. No response required.
TEP 7	Methods	[page 67]: Provide reference for interpretation of inconsistency factor plots, as these are not yet commonplace in network meta-analysis reporting.	We have added this as well as more description of the methods (in Appendix G).
TEP 7	Methods	[page 68]: Again, it would be useful to see the exact model structure used. Was this a covariates model with subgroup indicators as covariates?	We have added this information in Appendix G describing the analytical methods.
TEP 7	Methods	[page 68]: Authors mention precision evaluated "on the basis of sample size and, if size is adequate, the degree of certainty". I'm a little confused about the distinction, since degree of certainty (e.g. standard error) is a direct function of sample size.	We appreciate the confusion. We've clarified that the degree of certainty part is related to the magnitude of effect of the effect estimate and the limits of the CrI/CIs. In the full report, we removed the clause "and, if size is adequate, the degree of certainty", and added later when describing the methods for rating down in the other context, "When sample size was considered adequate, we further assessed precision based on the magnitude of the effects represented by the effect estimate and limits of the credible/confidence intervals. For outcomes where thresholds of clinically significant values were known…"
TEP 7	Methods	[page 68]: Are we talking about the outputs from random effects models here, because if so, effects corresponding to underpowered studies are automatically adjusted for strength of evidence via shrinkage toward the overall mean.	We have clarified on how our assessments on precision were in view of the magnitude of effect within the CrI/CIs – in relation to this domain's need to assess whether clinical decisions can be made from the results.

Commentator & Affiliation	Section	Comment	Response
TEP 7	Methods	[page 85]: Were these random effects meta-analyses?	All meta-analyses were random effects.
TEP 10	Methods	Yes to all. Level of detail was appropriate, but comprehensive.	Thanks for this comment. No response required.
Peer Reviewer 3	Results	95% credible intervals needs to be defined. I have not seen this terminology before.	Thanks. We have added a sentence for how people can interpret these, “Bayesian approaches provide variances using credible rather than confidence intervals, interpretable as the range of values within which there is a 95% chance of finding the true value of the effect.”
Peer Reviewer 3	Results	ES-11 (wording) SGAs probably decrease irritability and probably slightly decrease lethargy/social withdrawal...	Thanks for this comment. We have used both versions of decreased slightly (as per the reference cited for this narrative approach) and slightly decreased throughout the report.
Peer Reviewer 3	Results	ES-11: The statement about why the conclusion about stereotypic behavior is limited to acute treatment is not clear	After inclusion of a new study this finding is no longer applicable so we have not made any clarification here.
Peer Reviewer 3	Results	ES-13: The statement "Our meta analysis favored SGAs for hyperactivity" is unclear. Favored over what?	Thanks for this comment. This was risperidone over placebo and we have revised this in the ES and full report.
Peer Reviewer 4	Results	The results and key messages are exact and applicable. I am not aware of any literature that was omitted. The data represented seems fair. It would be beneficial to have real world data published including by race, socioeconomic class and location throughout the United States	Thanks for the comment. We have added an item in the research gaps section alluding to this comment.
Peer Reviewer 9	Results	The structure of the results is clear and important points are presented.	Thank you. No response required.
Peer Reviewer 9	Results	Goodness of randomization seems not to be part of ROB too. Given the typical small sample sizes within studies, imbalance between groups is likely to occur.	Baseline imbalance between groups was assessed as part of the “other” domain in the Cochrane risk of bias tool; we have added this description in the methods of the full report. Very few trials were at risk of bias from this.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 9	Results	Half of the studies did not control for important confounding variables – how this might affect the results?	This may have affected the results, although the within-study subgroup analyses overall did not indicate any particular systematic threats to differential effects. The confounding factors we considered during our assessments were not shown to modify the effects on harms, so it is questionable what effect this had. We added a note in the Limitations of this CER section, "Combining data from trials and observational studies for harms outcomes may have added heterogeneity to the results, although close inspection of the data plots (e.g., Figures 82-85) indicated high variability within both types of study design and no indication of a systematic bias in any direction. Our reports of within-study subgroup analysis and our meta-regressions attempted to help explain some of this variability."
Peer Reviewer 9	Results	Helpful to comment on the magnitude of the clinical and methodological similarities among the included studies within each condition.	We are not sure if this comment is requesting any changes, but we have reviewed our narrative (and in text tables) of the studies by condition and feel that they are quite comprehensive in terms of patient, clinical, and methodological factors.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 9	Results	In the text, the studies are referred by reference numbers. In the forest plots, the studies are referred by the authors. It would be very helpful to add reference numbers to the forest plot to have a complete understanding of the text. Also some numbers in the forest plot have many digits – keep the number of digits consistent and two is adequate.	We understand the importance of reader interpretation of the text. Because this was the only comment about difficulty following/understanding the text we don't think there's a large issue here. On further review we think the only confusion that may be important would be from not knowing which studies (when looking at the figures) were supporting observations on sub-group effects, or were removed for the various sensitivity analyses; we have added the study author names into the text for many of these sentences to help readers see the reasoning behind our analyses via the figures. There were several comments from reviewers about agreeing with our decisions throughout the results, although the added clarity may be helpful. Although we use multiple digits for the input values we calculated (e.g., SDs) all of our outputs are kept to 2 significant figures which is most important.
Peer Reviewer 9	Results	I would like to see the estimates of variance parameters and its 95% CrI in the forest plots.	We chose to report the I^2 values for this report, 1) it is more interpretable by our audience and 2) the between-studies variance parameter in a random effects Bayesian meta-analysis can be unstable and dependent upon the prior distribution (particularly for low number of studies, as many of our analyses were). We felt the I^2 statistic was better representation of the actual heterogeneity in the observed data.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 9	Results	If possible, comment on the clinical significance of the combined MD from meta-analysis.	We feel that our narratives describing the results in terms of strength of evidence (e.g. considering precision of effect estimates but also directness, study limitations etc.) and magnitude of effect in the tables and ES/discussion text is helpful for this purpose. A lack of defined thresholds for many outcomes restricts definitive conclusions in this respect; we have added comment that many of our inferences on magnitude of effect were based on clinical input.
Peer Reviewer 9	Results	Figures 8 – 12: data from the control group of Sikich 2004 and 2008 are used twice? The data from the control group are not the same in the same plot (Figure 10 to 12)? Since the data from the same generation are combined here, the data from the same study could be combined first to avoid use the same control data twice.	For these studies which were three-arm, we wanted to show and account for the results for both SGAs vs the FGA, therefore chose to separate the comparison in the analysis. We divided the sample size in the control arm by 2 to account for the lower precision from this approach but the effects are the same. We realized that 2 of the analyses did not split the sample size for one control group and we re-did these figures and analysis (Fig 9 & 10).
Peer Reviewer 9	Results	Figure 25: two studies, the combined estimate has wider CrI than the individual studies, no MA?	Thanks for pointing out this error. We have fixed the figure (the text was accurate without results from any meta-analysis).
Peer Reviewer 9	Results	Figure 52: Maybe OK not to combine.	We are comfortable keeping this meta-analysis (I2 41%); the results for each drug (main moderating variable of interest) would likely be very similar.
Peer Reviewer 9	Results	Weight and BMI are two variables highly influenced by length of followup. Is it valid to include all data? (Actually treatment duration showed to be an effect modifier later, if follow up time is the same as, or related to, the treatment duration). If data from different length of followup were analyzed in one MA, at least there should be a variable in the model to control for the differences.	We agree that for these outcomes the treatment duration may be a major factor. We have kept the combination of all durations, but added an adjusted analysis using this variable, and report on this within the section on subgroup analyses.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 9	Results	Figure G1 to Figure G4 should be in the text instead of appendix to help readers to have a better understanding of the available direct and indirect evidence, which in turn, helps to understand the validity and the results from the network MA.	We feel that the descriptions/tables (e.g., Table 2 and those in each section of the report) describe the (limited; biased for olanzapine) extent of indirect evidence quite well. We are aware of the length of the report as it stands and think readers, if interested, can easily view the appendix.
Peer Reviewer 9	Results	Figure 80 shows Molindone and Thioridazine ranks first and 3rd, however, based on Table 29, data for Molindone come from 20 patients, and Thioridazine, 15 patients. These results and such ranking are not reliable. How about treatment duration in these drugs? In this case, find the probability of "worst", which may provide more useful information.	We have commented on the imprecision for several drugs due to small samples, and have run an adjusted analysis for treatment duration. Thank you this suggested to find values for the worst; we have changed the analysis for this.
Peer Reviewer 9	Results	Results for inconsistency: please see the earlier comments in the Methods section.	We have enhanced are descriptions of the methods and results from this analysis in the text and appendix G.
Peer Reviewer 9	Results	In general, the direct and indirect evidence are sparse for most comparators. Focus on the few comparators with more data.	We had mentioned the limited precision from several drugs. We have added a comment about the most robust findings, "The relative harm from olanzapine is most robust compared with aripiprazole, quetiapine, and risperidone because of the precision in these estimates from larger sample sizes."
Peer Reviewer 9	Results	Figures 82 – 85 More informative to show data of the study level variables	These figures were added with the main purpose to show how results did not show any consistent differences between the different conditions (one of our subgroup variables). We have clarified this by removing mention of the figures from the other regression findings and left the reference within the text for effects by condition.
TEP 1	Results	Yes, adequate and appropriate.	Thanks for this comment. No response required.
TEP 1	Results	Clinical trial should be clearly defined as "randomized trials" because many consider also uncontrolled prospective studies as "clinical trials"	Thanks for pointing out this issue for clarity. We included both randomized and nonrandomized trials; our added definition of a nonrandomized trial will help clarify this and differentiate them from controlled observational studies.

Commentator & Affiliation	Section	Comment	Response
TEP 2	Results	Yes. The tables and figures are helpful.	No response required.
TEP 5	Results	I think the way in which the information is presented in the results is useful. Describing the study participants, then what the study looked at with the SOE and ROB is clear. Good to separate them..	We are glad these were useful. No response required.
TEP 5	Results	The Forest plot figures are not clear. In addition, the summary point, diamond, appears too large and the point estimates in the other studies are difficult to see	Thanks for the feedback. Because of limitations in the software we used for this report, we developed the figures manually. We have revised the figures to improve clarity. Please note that when accessed online, readers will be able to enlarge figures.
TEP 5	Results	Clear descriptions of the sensitivity analyses, why they were done and what they revealed throughout the report. It is very clear why studies were or were not included in the meta-analysis and why particular studies were removed for sensitivity analysis.	We are glad you approve of this. No response required.
TEP 5	Results	The within-study group effects sections are nice, concise, pertinent to important topics and clear.	We are glad these were useful. No response required.
TEP 5	Results	Good efforts to explain differences in outcomes if studies not equal in terms of interventions: page 146, line 3.	Thanks for the comment. No response required.
TEP 5	Results	Clear description of how harms were reported and differences as to study inclusion.	Thanks for the comment. No response required.
TEP 5	Results	All studies included appear appropriate.	Thanks for the comment. No response required.
TEP 6	Results	The text of the results is concise and appropriately augmented by the tables, figures, and appendices. I did not note any omissions and I felt the search was rigorous and complete (based on time parameters).	Thanks for the comment. No response required.
TEP 7	Results	[page 141]: Why is I2 reported? If a Bayesian random effects model was used, there is a direct estimate of heterogeneity automatically available (the variance of the random effects).	We preferred to use I^2 for two reasons: 1) it is more interpretable by our audience and 2) the between-studies variance parameter in a random effects Bayesian meta-analysis can be unstable and dependent upon the prior distribution (particularly for low number of studies, as many of our analyses were). We felt the I^2 statistic was better representation of the actual heterogeneity in the observed data.

Commentator & Affiliation	Section	Comment	Response
TEP 7	Results	[page 203]: Does the MCMC output merit 4 digits of precision? Does not seem to.	We agree and have made sure to include 2 digits in outputs.
TEP 10	Results	Figures were particularly helpful. Glad to see table summarizing studies that compared two SGAs. Can't think of any studies off hand that should have been included but aren't	Thanks for the comment. No response required.
Public Comment (Cynthia Russo, Janssen Pharmaceuticals, Inc.)	Results	[page 46]; Pali-low vs med vs high dose – There was a 6 point reduction onpositive symptoms PANSS score; p=0.003.	Our results (using data in published Table 2) are from comparisons between doses rather than between (all) the doses and the placebo.
Public Comment (Cynthia Russo, Janssen Pharmaceuticals, Inc.)	Results	[page 46]; Pali-low vs med vs high dse – When referring to the CGI-S and reduced illness severity, only the p value for the high does is presented (0.2); medium dose p<0.001.	Thank you we have added this.
Public Comment (Cynthia Russo, Janssen Pharmaceuticals, Inc.)	Results	[page 58]; SGAs versus placebo-Description of long term studies- risperidone vs placebo - There was a 3rd arm to this study92 that is not listed in the report: supportive therapy + placebo (n=28). The other two arms should be described as cognitive therapy + risperidone (n=43) and cognitive therapy + placebo (n=44)	Throughout the report we only describe the arms that were included for this review; the supportive therapy arm was not of interest.
Public Comment (Cynthia Russo, Janssen Pharmaceuticals, Inc.)	Results	[page 68]; SGAs Versus Placebo - "..one study included children ages 4 to 8. 110" The study methods include ages 3-7 and only the average age of patients in each group is available. The range of ages 4-8 is not presented in the publication.	Thank you we have corrected our error.

Commentator & Affiliation	Section	Comment	Response
Public Comment (Cynthia Russo, Janssen Pharmaceuticals, Inc.)	Results	[page 84]; Aripiprazole vs risperidone - The comparative outcomes of irritability, inappropriate speech, lethargy, social withdrawal, hyperactivity, and stereotypy are reported in reference, Ghanizadeh and colleagues.123	We realize the report may be confusing here. Our key points reflect the interpretation of the strength of evidence; such that "are not known" reflects insufficient strength of evidence rather than lack of evidence. The findings for these outcomes are described in the Detailed Analysis. We have clarified this within our section of the methods on interpretations, "We chose to use standard wording to describe how we interpreted the SOE and the magnitude of the effects for key outcomes;[61] our Key Points and tables of the strength of evidence (results chapter) and discussion relay these interpretations, while our Detailed Findings sections provide the exact fidnings regardless of their strength of evidence."
Public Comment (Cynthia Russo, Janssen Pharmaceuticals, Inc.)	Results	[page 113]; Eating disorder overview-Hagman[159] G1 Mean age is 16.2 ±2.5 years.	Thank you we have corrected our error.
Public Comment (Cynthia Russo, Janssen Pharmaceuticals, Inc.)	Results	[page 135]; Major AEs and major AEs limiting treatment- "Three RCTs [77,123,163] reported on numbers of patients discontinuing SGA treatment because of major AEs." The major AEs should be listed in parentheses as some patients may have discontinued but were not necessarily categorized as a major AE. For example, REF 163 does not appear later in the paragraph where the major AEs are reported. Ref 123 had one patient discontinue however it does not appear that this was considered a major AE as described later in this section.	We were not able to understand this comment. Our section on this page on Major AEs and major AEs limiting treatment is the only textual description of these AEs, so we are not sure what is meant by the wording "later in the paragraph". We only reported on data described by authors as major/serious AEs limiting treatment.

Commentator & Affiliation	Section	Comment	Response
Public Comment (Cynthia Russo, Janssen Pharmaceuticals, Inc.)	Results	[page 135]; Cardiac arrhythmias - "No patient receiving aripiprazole or risperidone (N = 60) had an abnormal ECG or pathological elongation in QTc values."[184] Note: Treatment with risperidone was associated with a slight increase of both mean QTc and QTd values (407.4 + 11.9 ms vs 411.2 + 13.0 ms, p< 0.05; and 40.0 + 4.4 ms vs 44.7 + 5.5 ms, p < 0.001, respectively).	Thank you for pointing this out. We only report on the proportions stated as having values in an "abnormal" or "pathological" range rather than any change.
Peer Reviewer 3	Discussion	The research gaps emphasize a number of important issues related to the medications, but given the relatively small to moderate effects sizes and substantial side effects, need to include some questions about how the decision is used to start medications and efficacy in related to behavioral interventions for irritable behavior in ASD, tics, and some other conditions	Thank you for this comment. We have added a point to this affect in the research gaps section, "Considering antipsychotics are recommended for use as adjunctive, or add-on, treatment for many conditions/symptoms, more studies examining these approaches (e.g., behavioral/family interventions with and without antipsychotics for hyperactivity or irritability) may help practitioners create guidance on when to start a trial of antipsychotics".
Peer Reviewer 4	Discussion	The major finding and limitations are states. I am unaware of any important literature that was inadvertently missed. Yes, this works is easily translated into new research.	We're glad you think so. We have revised and added to the research gaps section with suggestions from several peer-reviewers.
Peer Reviewer 9	Discussion	Another limitation of evidence is not much for head to head comparisons.	We agree and have added a sentence to this effect "In general, the small number of comparisons between different antipsychotics is a limitation in the evidence base." We also added a clause to the first research gaps about comparative studies). Our original conclusion also stressed this with "Overall, data for head-to-head comparisons (FGAs vs. SGAs, FGAs vs. FGAs, and SGAs vs. SGAs) were generally of insufficient or low SOE; therefore, few conclusions regarding the relative benefits and harms of different antipsychotics could be drawn."

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 9	Discussion	Provide some discussions on the validity, strength and limitations of the network meta-analysis.	We have added methods for our assessments and added a point about limitations to the discussion. “The findings from our network meta-analyses should also be considered exploratory in nature. Apart from the assumptions made for all meta-analyses, the network approach assumes transitivity, where we assume that all treatment nodes not present in any trial are missing at random, and there is nothing systematically different about the populations or interventions in the various trials. Because of these limitations we did not use these results for making our assessments of the strength of the body of evidence. We note, however, that the consistency between direct and indirect evidence was acceptable, and that the adjusted analysis factoring in treatment duration (shown as significant treatment modifier from the pairwise analyses) did not change the results.”
Peer Reviewer 9	Discussion	The conclusions could bring out more about the findings through the many many analyses.	We realize more could be added to the conclusions, although considering the report length and existence of an abstract and executive summary with more results we have only added a small amount. “For schizophrenia, there appears to be little or no difference between FGAs and SGAs for negative symptoms, positive symptoms, response rates, and global impressions of illness severity; deciding on which antipsychotic to use for this condition likely relies on close examination of the relative harms including considerations of their tolerance, management, and reversibility. The evidence examined suggests there may be little difference in effects between different doses of antipsychotics, although longer-term data would help clarify these findings.”

Commentator & Affiliation	Section	Comment	Response
TEP 1	Discussion	It is in general informative, but some statements could lead to misinterpretations by readers who are not entirely familiar with the methods and criteria of systematic reviews. a. In particular, the statement that "the majority of the trials had high risk of biases", while observational studies had only low or moderate risk of bias can lead someone to giving more weight to data from observational studies. This would be contrary to the fact that observational studies, by definition, are subject to more biases than randomized trial, for the basic reason that the treatment allocation is not randomly determined and there is no temporal control. In addition, NIH (and other agencies and journals) define clinical trial as any prospective treatment of human subjects with focus on health relevant outcomes. Thus, randomization is not a requirement for this definition of clinical trial. It may be appropriate therefore to specify in the abstract that the term "trials" here refers to "randomized clinical trials".	Thanks for this comment. We added a statement (results of methodological quality in ES and full report) to note that "the observational studies are still considered of poorer quality (e.g. ability to provide valid findings) than the RCTs, because of their inability to completely account for confounding by patient characteristics." We have clarified our definition of nonrandomized controlled trials to differentiate these from observational studies.
TEP 1	Discussion	b. Another issue is the statement that none of the evidence was rated "high strength of evidence". While this conclusion is derived by standard criteria for evidence-base medicine used in the Cochrane and other review groups, the evidence come from many controlled clinical trials conducted on antipsychotics in children. So, it may be useful to mention at least the 2 most common reasons for downgrading the findings.	We agree that this might lead to misinterpretations. We have added a couple sentences in the ES (implications) and final report (limitations) section, "The main reasons we downgraded the SOE was for risk of bias (largely from incomplete data due to study withdrawals) and imprecision from small samples or when the effects included possibility of substantial benefit or harm when insignificant findings were found (i.e., limiting confidence in findings of no difference). It should be recognized that attaining high SOE from trials of antipsychotics in children with psychiatric conditions is likely very difficult and the overall evidence reviewed should not be interpreted as lacking in credibility."

Commentator & Affiliation	Section	Comment	Response
TEP 2	Discussion	The major findings highlight the known need for additional research into both the effectiveness and the adverse event profile for both SGAs and FGAs in the child and youth population. This is clear from the review.	Thanks for the comment. No response required.
TEP 5	Discussion	Good summary for each condition of the major findings. The limitations are outlined and made clear. I don't know if all the important literature is included. The search was comprehensive and the discussion of the findings as they relate to other published findings suggests many/most relevant studies were found.	Thanks for the comment. No response required.
TEP 5	Discussion	Good description of the limitations of the CER. What and why were explained.	Thanks for the comment. No response required.
TEP 5	Discussion	Limitations of the evidence base is clear.	Thanks for the comment. No response required.
TEP 5	Discussion	Future research considerations nicely summarized the gaps prior to and after this CER was performed. The bulleted points could generate a number of future studies.	Thanks for the comment. No response required.
TEP 5	Discussion	It might be useful to list somewhere for the several outcomes it has been determined what constitutes a clinically meaningful difference. It is embedded in the narrative but with all of the numbers being reported, many of which are statistically significant the real importance of clinically meaningful difference gets lost.	We added a statement in the Limitations of the Evidence Base section, “There were few outcomes (e.g., tic severity, psychotic symptoms) for which we found clear evidence supporting a particular clinically important magnitude of effect; for most outcomes we relied on clinicians to help determine values for use in our assessments (e.g., >1 point change on the Clinical Global Impressions [CGI] scales, approximately a 10% mean difference for most measurement scales [10 points for scale of 1 to 100], RR values <0.75 for harm or >1.25 for benefit); effect sizes below these thresholds but having low or higher SOE for a difference were considered slight or small.
TEP 6	Discussion	Perhaps the future research section was a bit truncated but did mention all the main categories of outstanding issues. Limitation section was excellent	Thanks for the comment. We have revised and added to the research gaps section with suggestions from several peer-reviewers.

Commentator & Affiliation	Section	Comment	Response
TEP 6	Discussion	At some point in the report, some mention/summary of the differences between this revision and the original report could be made.	Appendix A outlines the differences between the reviews in terms of the inclusion/exclusion criteria and methods. Because there were multiple changes in terms of outcome definition (most outcomes were changed in some manner to add specificity) it is very hard to compare the reviews in terms of findings.
TEP 7	Discussion	Discussions and conclusions are clear and adequately described. I am not in a position to critique the literature citations.	Thanks for the comment. No response required.
TEP 10	Discussion	Clear	Thanks for the comment. No response required.

Commentator & Affiliation	Section	Comment	Response
Public Comment (Cynthia Russo, Janssen Pharmaceuticals, Inc.)	Discussion	As you point out, a 3-week placebo-controlled RCT compared three doses (2.5, 5, and 10 mg twice daily) of asenapine. All three doses offered significant improvement over placebo for manic symptoms, response rates, and global impressions of severity and functioning. The results suggest a dose-response relationship for the outcomes of manic symptoms and response rates (both related to YMRS scores; p values 0.5, 0.001, and 0.0001, respectively), although not for depression or for global impressions of severity or functioning. Only the 10 mg twice daily group was favored over placebo for depression scores on the CDRS. A 50-week open-label extension study was included as an option to patients included in the 3-week efficacy study. The objectives of this study were to collect safety data in long term treatment of pediatric subjects with a manic or mixed episode associated with bipolar I disorder (primary) and to collect exploratory long-term efficacy data of asenapine in these patients. [Data on file. Actavis. Publication pending] Somnolence and Sedation were the most commonly reported Treatment Emergent Adverse Events (TEAEs), and may be treatment-limiting in a minority of subjects. These TEAEs combined with dizziness and oral hypoesthesia occurred more frequently in asenapine-na ve subjects and may, thereby, be TEAEs to consider with asenapine treatment initiation in this population. [Data on file. Actavis. Publication pending.]	We are happy our interpretations were correct for the controlled phase. Thank you for sharing the results from the open label phase, although this does not meet our inclusion criteria.

Commentator & Affiliation	Section	Comment	Response
Public Comment (Cynthia Russo, Janssen Pharmaceuticals, Inc.) (continued)	Discussion (continued)	A total of 25.5% of subjects in the 50-week study experienced clinically significant weight increase (i.e., 7% increase in body weight at endpoint), which is not necessarily unexpected in a developing pediatric population. Overall mean (SD) weight gain at study endpoint was 3.0 (5.0) kg. Weight gain based on weight percentile (i.e., adjusted for growth, as based on age and sex) continued to increase after treatment initiation in this extension trial, which is also consistent with alterations observed in fasting metabolic chemistry parameters in a minority of subjects at study endpoint. With respect to meeting criteria for the metabolic syndrome (MBS), results suggest that asenapine treatment may be associated with development of MBS in a small minority of subjects and that this is reversible in some individuals over time. [Data on file. Actavis. Publication pending.] Flexibly dosed (2.5 mg, 5.0 mg, and 10.0 mg BID) asenapine was generally safe and well tolerated in pediatric subjects with bipolar disorder treated for up to 52 weeks. The majority of subjects were administered asenapine 5.0mg BID or 10.0 mg BID.	
Peer Reviewer 3	Clarity & Usability	Generally well structured. A few areas where clarity could be improve are mentioned above.	Thanks for the comment and suggestions.
Peer Reviewer 4	Clarity & Usability	This is an organized work with its main points presented and conclusions relevant to practice decisions. It contributes further understanding, but it doesn't appear to be new information. I submitted a few edits on the attached document and hope it brings more clarity to this qualitative work. Thank you for the opportunity to review!	Thanks for the comment and suggestions.
TEP 1	Clarity & Usability	Yes, the report is well organized. Conclusions are consistent with stated methodological premises.	Thanks for the comment. No response required.
TEP 2	Clarity & Usability	Yes	Thanks for the comment. No response required.

Commentator & Affiliation	Section	Comment	Response
TEP 5	Clarity & Usability	Fantastic report. Dense and weighty but there was a lot of information to sort through. It is well organized, logical and consistent in its structure. The summary sections are succinct and to the point. They are quickly read and digested. The lack of evidence is concerning. Particularly with the marked increase in use over the last several years in very young children.	Thanks for the comment. No response required.
TEP 5	Clarity & Usability	From a Medicaid policy perspective, this report provides a number of launching points for guideline development and analysis of current practice.	Thanks for the comment. No response required.
TEP 6	Clarity & Usability	Very clear and concise. Reasonably well-organized.	Thanks for the comment. No response required.
TEP 7	Clarity & Usability	The report is very well structured and easy to read and understand. The authors do an outstanding job of summarizing the evidence, both informally and formally (where meta-analysis was applied).	Thanks for the comment. No response required.
TEP 10	Clarity & Usability	Clear and easy to navigate. Useful information.	Thanks for the comment. No response required.