Comparative Effectiveness Research Review Disposition of Comments Report

Research Review Title: *Pharmacologic Therapies for the Management of Crohn's Disease: Comparative Effectiveness*

Draft review available for public comment from September 7, 2011 to October 5, 2011.

Research Review Citation: Hutfless S, Almashat, S, Berger Z, Wilson LM, Bonson E, Chen Q, Donath E, Herlong F, Puhan MA, Selvaraj S, Tuskey A, Vasilescu A, Bass EB, Worthington M, Palamaner G, Lazarev M. Pharmacologic Therapies for the Management of Crohn's Disease: Comparative Effectiveness. Comparative Effectiveness Review No. 131. (Prepared by Johns Hopkins Evidence-based Practice Center under Contract No. 290-2007-10061-I.) AHRQ Publication No. 14-EHC012-EF. Rockville, MD: Agency for Healthcare Research and Quality; February 2014. www.effectivehealthcare.ahrq.gov/reports/final.cfm.

Comments to Research Review

The Effective Health Care (EHC) Program encourages the public to participate in the development of its research projects. Each comparative effectiveness research review is posted to the EHC Program Web site in draft form for public comment for a 4-week period. Comments can be submitted via the EHC Program Web site, mail or email. At the conclusion of the public comment period, authors use the commentators' submissions and comments to revise the draft comparative effectiveness research review.

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

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

Source: http://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=1869
Published Online: February 25, 2014

Commentator & Affiliation	Section	Comment	Response
TEP-1	General comment	The conceptual model needs work	We have provided a new figure in the introduction, and we have revised the text throughout to focus on remission rather than on response to treatment.
TEP-1	General comment	The document never defines “induction” or “maintenance” but we can glean from the tables that the following historical concepts are being used: 2-4 weeks “immediate response”; 8-12 weeks “induction”; 15-19 weeks “4 months”; 18-24 weeks “long-term”; 22-30 weeks “six months”; 48-54 weeks “1 year”. The authors should consider the possibility that these terms are anachronisms. The report does not provide a clear operational definition of maintenance that is biologically based. The clinical guidelines have organized themselves as induction and maintenance. It is a historical reality that the GI docs conceptualized treatment in this manner, but it may be an artificial construct. Perhaps the historical conceptual model cannot easily be modified, but really, this is the same outcome at two different time-points. And the review is right to show not merely two time-points, but six. For many other disease (e.g., RA), the questions are: is there an initial response? Is the initial response lost? These are called primary and secondary treatment failure.	We changed the time periods of interest for induction and maintenance of remission to be more consistent with how the terms have been used in the literature on treatment of Crohn’s disease. We agree that these terms are anachronisms, but we think it is important to be consistent with how the terms have been used in this field. We have diagrammed and defined induction and maintenance of remission in the introduction.
TEP-1	Executive Summary	Thus, ES-9, first full paragraph, “The results…” is long and complicated, pointing out the fundamental challenge of the review.	In the methods section of the Executive Summary (especially in the section on Data Synthesis and Meta-Analysis), we have tried to convey how we organized our approach to this complicated review.
TEP-1	Executive Summary, Tables B,C,D	And, Table B has two identical sections for KQ1 and 2. Further, Tables C and D are inter-weaving and repetitive with respect to timepoints, with Table C skipping over weeks 15-19, and both tables presenting information beyond weeks 18. Why is 18-24 at the far right?	We have removed Tables C and D, and replaced them with Figures C, D, E and Table F. Figures C and E focus on placebo-controlled trials for induction and maintenance of remission, respectively, and Figure D focuses on head-to-head trials for induction of remission. We have revised the timepoints listed in Table B.

Commentator & Affiliation	Section	Comment	Response
TEP-1	General comment	I'm suggesting it might simplify matters to re-organize the entire document by drug, using the top-down framework. The document would then address "what does this drug accomplish over the course of time" instead of "which drug should we use for relief at timepoint X" followed by a rather repetitive section focused on "which drug should we use for relief at timepoint Y".	Final report order was decided after discussion with the Task Order Officer and Associate Editor. In the first part of the Results chapter, we describe the organization for the Results chapter. We reported the results for adults and children separately: "We reported the results of our systematic review first according to KQ. For the efficacy results, we organized the results for each KQ by medication class… Within each medication class, we reported the study design and population characteristics, the key points, strength of evidence (SOE) grading, and then the outcomes results… The outcomes results were arranged by comparison. We present first the monotherapy placebo-controlled trials, followed by monotherapy head-to-head comparisons, then combination therapy placebo-controlled trials, combination therapy versus monotherapy comparisons, and then combination therapy head-to-head comparisons… We organized the safety results by outcome then by medication class. We report the subgroup analyses at the end of each KQ."

Commentator & Affiliation	Section	Comment	Response
TEP-1	General comment	Further, I found that separating the findings for adults, children, and subgroups was harder, not easier. So: Anti-TNF agents; Agent 1 monotherapy: adults, kids, subgroups.; Agent 1 combination therapy: ditto.; Etc.; KQ 1 would then be "What is the effectiveness (and I would stick safety here too) of Anti-TNF Agents" with these outcomes having a time dimension. I do agree that the surgery question is distinct.	Final report order was decided upon after discussion with the Task Order Officer and Associate Editor. In the first part of the Results chapter, we describe the organization for the Results chapter. We reported the results for adults and children separately: "We reported the results of our systematic review first according to KQ. For the efficacy results, we organized the results for each KQ by medication class… Within each medication class, we reported the study design and population characteristics, the key points, strength of evidence (SOE) grading, and then the outcomes results… The outcomes results were arranged by comparison. We present first the monotherapy placebo-controlled trials, followed by monotherapy head-to-head comparisons, then combination therapy placebo-controlled trials, combination therapy versus monotherapy comparisons, and then combination therapy head-to-head comparisons… We organized the safety results by outcome then by medication class. We report the subgroup analyses at the end of each KQ."
TEP-1	Executive Summary	The remainder of the ES is terrific	Thank you.
TEP-1	Introduction	Comments per conceptual model, otherwise it's good	Thank you for reviewing our evidence report!

Commentator & Affiliation	Section	Comment	Response
TEP-1	Executive Summary, Methods	Page ES-4, line 27, drop “as pertains to patient-reported outcomes”.	The reviewer is referring to the wording for KQ 4: “KQ4: What is the comparative effectiveness of agents used to prevent post-operative recurrence in Crohn’s disease as pertains to patient-reported outcomes?” Thank you for the suggestion. However, we chose not to drop this part of KQ wording.
TEP-1	Executive Summary, Methods	Page ES-5, Table B lists KQs but does not attach each to a descriptive phrase.	A descriptive phrase of the Key Questions is included as a footnote in the table.
TEP-1	Results	Comments per conceptual model, otherwise it’s complete and good	Thank you.
TEP-1	Discussion/ conclusion	Excellent	Thank you for reviewing our evidence report!
TEP-1	Clarity and usability	Objectives, please add “induction or maintenance of remission in placebo-controlled or head-to-head randomized controlled trials (RCTs),” and then this phrase can be shortened (perhaps) in line 16.	We have edited the objectives to state that we are comparing the efficacy and safety of therapies to reflect that we are only including RCTs for measures of efficacy.
TEP-1	Clarity and usability	Line 33, “specific biologics were more effective”…. Is vague. This would not be a problem, except that line 41 says “other… measures of efficacy”. Please re-word to reduce confusion. Considering that the line 41 reference is quite comprehensive, I suggest you change it to the general.	We have removed this sentence from the abstract and replaced it with two more specific sentences.
TEP-1	Clarity and usability	Line 46, calling out PML is a little from left field; not mentioned in line 26.	The reference to PML was removed from the abstract.
TEP-1	Clarity and usability	Line 49, the conclusion is not tightly worded. Efficacy of those agents against placebo? And by treatment options” in line 53, do you mean that they have not been subjected to head-to-head comparisons? This term was used in line 17 and could appropriately be used in the conclusion.	We have revised the conclusions to be clearer. We added a statement that most of the evidence for efficacy comes from placebo-controlled trials.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer-1	General comments	Hufless et al have conducted an extensive study of effectiveness of therapies to treat Crohn's disease (CD). Main findings do not vary substantially from other reviews on this topic. This review only contains randomized, controlled trials (RCT) for issues of induction and maintenance and patient centered outcomes, and observational studies along with RCT for safety. The end point of interest was remission. Induction endpoints were examined at 2-4 weeks, 8-12, 22-30, 48-54 and 18-24 months. For maintenance, 15-19 weeks, 22-230, 48-54 and 8-24 months. 184 studies were eventually included, after examination of 23,286 found on a literature search. For KQ1, of 53 trials, only 27% were of "good" quality, 48% were fair. For KQ2, of 46 trials, 24% were good and 52% fair. For KQ3, "many" studies were of poor quality. For KQ4, only 2 studies were included. It is unclear who the target audience is for this review, I assume those clinicians who care for patients with Crohn's disease. The target population appears to be patients with a range of disease activities.	Thank you for taking the time to review our draft report! We added wording to the statement of the purpose of the report (ES-3) to clarify that the target audience is clinicians involved in the care of patients with Crohn's disease, as indicated by the reviewer. In the analytic framework, we indicate that the target population is patients with active Crohn's disease for KQ1 and KQ3, patients in remission for KQ2 and KQ3, and post-surgery patients for KQ4.
Peer Reviewer-1	Appendix G	Appendix G is very confusing and not typical of the types of charts that most people read, the columns are not parallel and hard to follow.	Appendix G has been revised to make the columns easier to follow. It is now organized by sub-population.
Peer Reviewer-1	Introduction	Adequate and no issues	Thank you for reviewing our evidence report!
Peer Reviewer-1	Methods	I am surprised that the authors chose to examine ustekinumab, an agent that is not FDA approved and currently is in Phase III trials. Literature is for the Phase II trials and perhaps the rationale is that this is an agent available off label. However, that reasoning is valid for antibiotics, a class of therapy that is not included at all and the authors need to explain why they excluded these studies. The authors were very strict in the inclusion criteria. Those studies for maintenance that did not have a washout period were excluded. In the real world patients get switched to new therapies when the old drug is still on board. Using remission as an endpoint is fine, but it is a goal and in some quite lofty. I am not sure that the conclusions can be taken on face value with that caveat.	Ustekinumab was removed from report. The comment regarding the wash out period only applies to crossover studies. We excluded the one crossover study evaluating ustekinumab, so this point is moot. We agreed in consultation with AHRQ to use remission as an endpoint. We comment on remission and concomitant therapies in the discussion.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer-1	Methods; quality scores	As far as the scoring of the quality of the studies, studies conducted back in the 1970's were included. It is well known that the science of trial design has come a long way and that quality from a study done in the 1970's would not have anywhere near the quality of one performed just two years ago. The authors did not contact any pharma companies for data that may not have been published. They admit to a publication bias but also may change their conclusion is that data were available to them. "Indirect comparisons" of natalizumab versus anti-TNF therapies and comparisons between anti-TNF agents just because of similar mechanism of action is too simplistic and misleading.	We removed the indirect comparison after discussion with AHRQ's Task Order Officer. Pharmaceutical companies were not contacted for data. We mention publication bias as a limitation because it was detected in some of the meta-analyses we tried to conduct.
Peer Reviewer-1	Results	Clinicians are not going to understand "strength of evidence is low to demonstrate no difference between…" Does that mean there is a difference or that there is insufficient evidence to detect a difference"?	We have reworded the Key Points so that the conclusion is stated first, then the strength of evidence.
Peer Reviewer-1	Results	The 5-ASA meta analysis was feasible but only included 4 studies, it was statistically significant at weeks 48-52. These results, as pointed out, are at odds with the conclusions of the 14 studies overall. I think this is confusing and some further discussion of these findings should be included as there are many practitioners who treat patients with Pentasa and may mis-interpret your findings.	Agreed, this is distracting so we removed the 5-ASA meta-analysis, especially given that the largest study in the meta-analysis was for olsalazine, which is not used in practice
Peer Reviewer-1	Results	I have an issue with the studies that you excluded. Just a random sampling of articles listed as "No Crohn's patients in the trial" seems to be used as rationale quite indiscriminantly. For example Agnhott 2003 there were 26 patients with CD studied; Rosh JR et al they report on 115 CD pediatric patients and Sandborn WJ 2010 for the certolizumab trial Crohn's disease is even in the title! These studies may have been excluded but obviously for other reasons than "No Crohn's disease patients included" In KQ3, 44 RCT and 46 observational studies are included, but in the Methods section it says 45 RCT included and in Table 15 it appears to be 43 distinct studies and 48 observational in Table 16.	The exclusion report has been updated. Studies that were excluded for "No Crohn's patients" were generally excluded because they did not include an exclusive Crohn's population. The wording has been modified to say "RCT patient population not exclusively Crohn's or observational study not exclusively IBD."
Peer Reviewer-1	Discussion/ conclusion	How are we ever going to do long term, large studies without pharma/industry funding?	In the Future Research Needs section, we have revised this to suggest how we could minimize conflicts of interest, such as having trials that are independent of industry input in the design, analysis, interpretation, and design to publish.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer-1	Discussion/ conclusion; KQ4	Not sure you get anything meaningful out of KQ4, maybe just a discussion of minimal information on this outcome would be adequate. Another difference between this review and the ACG review is that the ACG review included antibiotics. Since there is no standard for a clinically meaningful absolute difference, it is difficult to make conclusions about the results of meta analyses and systematic reviews, the authors do point this out but perhaps also one of the Key Points in the Executive Summary should be how heterogeneous the data really are and that trying to be rigorous and only examine well done RCT will not answer the question(s) at hand.	a) KQ4 was chosen because it had not previously been addressed in prior systematic reviews. We agree that little can be concluded other than that post-operative patient-reported outcomes have not been well studied. b) We mention in the discussion that an additional difference in our review from the ACG review was our decision not to include antibiotics. c) We emphasize the study heterogeneity in the executive summary.
Peer Reviewer-1	Discussion/ conclusion; future research needs	The authors appropriately recognize Future Research needs and opportunities to address are the creation of national registries (the CCFA is actually doing this now, with PatientsAsPartners) updated guidelines (the ACG just last year came out with its latest guideline and the European Collaboration just published theirs) and conducting large RCT to address long term effectiveness (not very likely in this current climate).	We searched clinicaltrials.gov for additional trials and results. We did not find any results that differed from those in the manuscripts as the clinicaltrials.gov results were updated after the publications.
Peer Reviewer-1	Clarity and usability	The report is well structured and organized, it is easy to follow. However, I do not think that the conclusions from this report can be used for anything but discussion for where the data are lacking and how future trials should be designed. The inclusion criteria were very strict, and as discussed above the conclusions are not written for a clinician but rather a statistician or Methodology guru.	We have included a section in the Discussion called a “Clinical Perspective.”
Peer Reviewer-2	Abstract	The abstract should note the limited number of trials directly comparing different treatment strategies.	Added “Twenty-three percent of trials directly compared different treatment strategies.”
Peer Reviewer-2	General comments	This is an important compendium of the clinical trials that have been performed in Crohn’s disease. It obviously required an enormous amount of time and effort to produce this document. It will in general be a valuable resource.	Thank you for taking the time to review our draft report!
Peer Reviewer-2	Introduction	I am not sure that the mechanisms of action listed in Table A are all correct. Why are azathioprine and mercaptopurine listed as having different mechanisms of action? Should methotrexate be listed as inhibitor of folate metabolism which leads to increase in adenosine concentration?	Thank you! The mechanisms of action in Tables A and 1 have been reviewed and corrected.
Peer Reviewer-2	Methods, general	The methods were generally adequately described although the application of the methods to determine the strength of evidence was obscure.	We added extra details about how we rated the strength of the evidence.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer-2	Methods, purpose	The stated purpose of the review was to determine the comparative effectiveness and comparative safety of therapies used to induce and maintain remission of CD. Yet the inclusion criteria and studies available for review meant that the authors generally looked only at efficacy in RCTs and these were mostly placebo controlled RCTs.	We added to the limitations section of the discussion that there was a lack of head-to-head trials.
Peer Reviewer-2	Methods, Strength of evidence	Perhaps I missed it, but I could not tell what criteria led to a categorization that a comparison "favors neither" rather than "favoring specified drug". Generally these were in the setting of low SOE.	In the Data Analysis and Synthesis section of the Methods, we added what we considered to be a clinically meaningful difference. This was used to determine "favors neither" or "favors specified drug."
Peer Reviewer-2	Results	What does it mean that "prednisolone was favored over budesonide for infections."? Does this mean that prednisolone causes more infections or fewer infections?	The term "favored" is clarified in the text of the Executive Summary to mean fewer adverse events.
Peer Reviewer-2	Results	I struggled to see how the authors came to their strength of evidence conclusions. For example, in Table D, Natalizumab vs. Placebo has 1 trial with 339 people, 25% RD and highly statistically significant. This was rated as Low SOE. Certolizumab vs. placebo had 1 trial with 428 patients with RD 19% and the same RR=1.7 yet is rated as High SOE. Do we really think that there is moderate strength of evidence that infliximab is not superior to placebo for rapid induction of remission based on 1 study with RD of 23%, RR 6.4; CI 0.9-45.3? (Table C) These are just 2 examples, but it calls into question the entire projects rating of the SOE.	We have moved the tables showing how we graded the strength of evidence into the main body of the report. We also added more detail about how we graded the strength of the evidence in the Methods chapter. We graded the strength of evidence based on the risk of bias, consistency, directness, and precision of the results, not the number of studies and the effect size. We have reviewed and revised our strength of evidence grading so that it is consistent with our Methods.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer-2	Results	"The strength of evidence was low that the combination of infliximab and azathioprine was more effective than corticosteroids alone in inducing a steroid-free remission at weeks 26 and 52 (absolute RD across time points, 19% to 24%; corticosteroid rate, 36% to 42%)." The authors have misinterpreted this study as the comparison was to steroids with thiopurine or anti-TNF salvage therapy if needed. So this was not just infliximab and azathioprine vs. steroids.	We appreciate the difficulties of classifying this study. We took efforts to describe the protocol of this study in the text (see the Results chapter, in the TNF-Alpha Inhibitors section, under the Infliximab and Thiopurine Versus Corticosteroids Alone sub-section). We chose to include this study as it was the only one to directly compare step up versus top down.
Peer Reviewer-2	Results	I continued to struggle with the interpretation of the natalizumab data. In table 4, the authors basically conclude that natalizumab is ineffective for induction with moderate to high SOE. Yet as noted by the authors on page 29 the drug was better than placebo at weeks 2-4 and also at weeks 8-12 if you don't include the one trial with only a single dose of the drug.	Table 5 (previously Table 4) focuses on the last available time point. The largest study (Sandborn 2005) did not show a clinically significant difference at the later time points. Additionally, we were not able to perform a meta-analysis at the later time points because of design heterogeneity. We do have a footnote in the table stating that natalizumab was more effective than placebo at 2-4 weeks with high strength of evidence. This is also noted in the Key Points.
Peer Reviewer-2	Results	Why was the strength of evidence low for budesonide to maintain remission when the pooled OR from meta-analysis of 4 studies gave OR=1.0? Seems like high SOE. Of course this must also be contrasted to the conclusion that SOE is moderate that budesonide is superior to mesalamine to maintain remission. It was not possible for me to find every area of inconsistency such as this, but the examples I have provided lead me to question the basis of these conclusions.	We re-evaluated the strength of evidence for budesonide vs. placebo to maintain remission. We graded the strength of evidence to be low because the risk of bias was medium, and the results were inconsistent and imprecise. Budesonide versus mesalamine was graded as moderate because the risk of bias was medium, but the results were precise.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer-2	Results	The authors should read several letters to the editor in Gut about the Fidder 2009 study on mortality and cancer. The Fidder study was methodologically flawed by immortal time bias. With the correct analysis, very different results were obtained.	We appreciate the reviewer's point, and we have added text to acknowledge that the study was flawed by immortal time bias, as explained in detail in the letter to the editor.
Peer Reviewer-2	Results	I was surprised that the authors excluded studies of cancer outcomes when the control group was the expected rate in the general population. See meta-analysis by Kandiel for association of thiopurine therapy for IBD with lymphoma.	Our inclusion and exclusion criteria were uniformly applied. Table 4 states that patients with any inflammatory bowel disease were included as a sensitivity analysis. Studies that compared Crohn's disease patients to patients without inflammatory bowel disease were excluded.
Peer Reviewer-2	Results	Perhaps greater emphasis in the results or the discussion section could have been placed on trials that compared different strategies as opposed to placebo controlled trials of adding a new therapy on to other therapies. These are the trials that get us closest to true comparative effectiveness.	We re-organized the results after discussion with AHRQ's Task Order Officer. However, we present the placebo-controlled trials first because this establishes efficacy. We have created headers throughout the results chapter to orient readers to placebo-controlled vs. head-to-head trials.
Peer Reviewer-2	Discussion/ conclusion	The discussion and conclusions are likely appropriate, but do not offer any particular new insight. Before this systematic review was undertaken, we knew that there were very few head to head comparisons of different treatment strategies for CD. The authors of this review have confirmed this.	We feel it is critical to systematically define what is out there. This helps us to better understand the gaps in the literature, and prioritize for the future.
Peer Reviewer-2	Discussion/ conclusion	The document would be strengthened by specifying what evidence would be needed to move some of the key questions from low level of evidence to high level of evidence.	Thank you very much for suggesting this. We are creating a separate report on Future Research Needs, which will list prioritized future research needs and recommend study designs to address them.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer-2	Discussion/ conclusion	The authors recommend that the Crohn's disease community develop evidence-based treatment algorithms and guidelines based on the best available evidence, and updating the guidelines regularly as new evidence becomes available. Yet, the review suggests that there are very few areas where there is even moderate level of evidence to come to a conclusion.	We have revised the Future Research Needs section of the Discussion so that this statement is no longer there.
Peer Reviewer-2	Discussion/ conclusion; Future Research Needs	In the section on future research needs, the authors again emphasize the need for evidence on the step-up versus top down approach to treatment. I think that this should be expanded upon to recognize that this is but one possible variation of treatment strategies. There are others that require further study. For example, the recent SONIC trial suggests a therapeutic advantage for some patients using combination anti-TNF and thiopurines vs. either agent alone. However, the COMMIT trial did not find a benefit from combination anti-TNF and methotrexate. Furthermore, there are limited data on sequential monotherapy (e.g. anti-TNF followed by thiopurines or methotrexate if anti-TNF therapy does not produce remission). Thus, there remain several important areas for comparative effectiveness research that has not been adequately addressed in prior studies.	For efficacy endpoints, we only included randomized controlled studies. The COMMIT study (comparing combination infliximab/methotrexate versus infliximab alone), most pertinently, was not published at our last search (June 2011), and still has yet to be published as a manuscript. Thus, we did not feel we could fairly include this study for comparison to SONIC (which is well covered in our report).
Peer Reviewer-2	Discussion/ conclusion; Future Research Needs and safety outcomes	The authors recommend that safety outcomes be specified in advance and included as primary or secondary outcomes in RCTs. This implies that the important adverse event is hypothesized in advance. Often, we are not aware of the potentially important adverse event until after the trial is completed. PML with natalizumab is a perfect example. Furthermore, sample sizes for studying rare adverse events are typically much larger than for studying efficacy. This is why most safety data come from post marketing observational studies. So while this sounds like a prudent recommendation, it is often not feasible.	We have revised the Future Research Needs section of the Discussion. We now state, "Researchers need to design safety studies so they have sufficient power to detect clinically meaningful differences between medications."
Peer Reviewer-2	Discussion/ conclusion; Future Research Needs	The authors recommend that future research needs include "…examining the reasons for imbalance of potential confounders in the setting of RCTs…" The reason for this is known. It is bad luck. The point of randomization is to balance both known and unknown confounders between the treatment groups. But it does not always work. The larger the study the more likely it will work. Regardless, any difference between treatment groups in an RCT is by definition due to chance.	This has been removed from the Discussion.
Peer Reviewer-2	Clarity and usability	Some of descriptions are difficult to understand. For example, "the strength of evidence was low that there was no difference…" The authors understandably had to select specific time points because different studies use different time points for measuring the primary outcome. However, it results in limited ability to use clinical and epidemiologic judgment to draw conclusions with greater strength of evidence.	We have modified our summaries of the strength of evidence on each comparison to include data from what we thought were the most important time points. We tried our best to balance the desire to show all available data with the desire to avoid overwhelming readers with too much data.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer-2	Clarity and usability	Perhaps greater emphasis in the results or the discussion section could have been placed on trials that compared different strategies as opposed to placebo controlled trials of adding a new therapy on to other therapies. These are the trials that get us closest to true comparative effectiveness.	We agree with the reviewer that the emphasis should be on head-to-head trials. Unfortunately, we found a limited number of head-to-head trials that met our inclusion criteria. Where possible, we have tried to highlight the head-to-head trials. In the Executive Summary we have a figure that summarizes head-to-head trials of biologics (Figure D). Throughout the results section for KQ1 and KQ2, we have headers orienting readers to the head-to-head trials. After discussion with AHRQ's Task Order Officer, we have ordered the results of the report as (a) monotherapy vs. placebo; (b) monotherapy vs. monotherapy; (c) combination therapy vs. placebo; (d) combination therapy vs. monotherapy; and (e) combination therapy vs. combination therapy. We have included headings in KQ1 and KQ2 to orient the reader.
Public: UCB	Results	For key question one, the objective was to evaluate the comparative effectiveness of therapies alone or in combination used to induce remission in adults with moderate-to-severe Crohn's disease. In this review, three clinical trials were used to evaluate certolizumab pegol (CZP) – two phase II dose-response studies, and one phase III efficacy trial. Please note in the review that CZP is not indicated for the induction of remission of Crohn's disease. It should also be mentioned that the doses evaluated in the two phase II studies were not the currently approved doses for CZP in the treatment of Crohn's disease, therefore, would limit the finding's usefulness for clinicians. As stated in the approved Prescribing Information, the initial dose of CZP is 400 mg subcutaneously administered at weeks 0, 2, and 4, followed by maintenance dosing of 400 mg every four weeks thereafter in patients who respond. This dose was evaluated in the phase III efficacy trial for Crohn's disease.	We didn't limit our search to the dose used in practice. We considered all drugs for induction and maintenance, regardless of how they are indicated. We tried to include dosing information in our results whenever we could.

Commentator & Affiliation	Section	Comment	Response
Public: UCB	Executive Summary, KQ1	We would also suggest including in the Executive Summary in Table A, page ES-2, the indicated dosage from the Prescribing Information of the treatments listed. For those treatments that are not indicated for Crohn's disease, the recommended doses evaluated in clinical trials or stated in clinical practice guidelines could be included. Finally, we request consideration that the review limit its evaluation to only the phase III efficacy trial of CIMZIA for the induction of remission and exclude the two phase II trials, as these studies were not designed for that endpoint. It should be noted that CZP is not currently available as an intravenous formulation	The phase 2 trials have met our inclusion criteria, as well as the study using intravenous formulation. We have noted that there are applicability issues with some doses, and the intravenous route. We list what the current approved route is for each drug.
Public: UCB	Results, KQ1	For patient reported outcomes on page 20, the report concludes that "the strength of evidence was high to demonstrate no difference comparing CZP with placebo in improving patient-reported outcomes at week 12 and 26 (absolute between-group difference in change in mean IBDQ from baseline across time points, 5 to 11; placebo change in IBDQ, 18 to 21)." Per PRECiSE 1, IBDQ response was defined as an increase of at least 16 points in total score compared with baseline (week 1). At week 26, more patients treated with CZP vs. placebo had an IBDQ response (42% [140/331] vs. 33% [108/328]; *P*=0.01), with a mean increase from baseline of 26.4±35.1 points vs. 20.5±33.1 points, respectively (*P*=0.03). The report conclusions should take into consideration the significant patient-reported outcomes findings of the PRECiSE 1 trial.	We added to our Methods chapter under the Data Analysis and Synthesis section that we considered a 17-point difference in the Inflammatory Bowel Disease Questionnaire to be a clinically meaningful difference. We chose the 17-point difference based on the manuscript, Irvine EJ. Development and subsequent refinement of the inflammatory bowel disease questionnaire: a quality-of-life instrument for adult patients with inflammatory bowel disease. J Pediatr Gastroenterol Nutr. 1999;28(4):S23-7. The between-group difference in mean IBDQ from baseline did not meet this difference.
Public: UCB	Results, KQ2	Crohn's disease is a chronic inflammatory bowel disease with a relapsing and remitting course in 73% of patients.[1] Induction and maintenance of remission are two important goals of therapy. For the maintenance of remission, the draft review included the time point range of 18–24 months to evaluate long-term maintenance of remission; however, due to exclusion of studies that did not have a comparison group, data from open-label extension studies were excluded.	Correct, we only used studies with a comparison group.
Public: UCB	Results, KQ3	For the evaluation of long-term remission, it is difficult to consent patients to long-term randomized clinical trials resulting in randomization of patients to placebo, with the prospect of leaving their disease untreated. Therefore, we would like the reviewers to reconsider inclusion of data published in peer-reviewed journals from open-label extension studies to evaluate long-term time points. There is value in evaluating data from these long-term trials, both from the safety and efficacy perspectives. Specifically regarding CIMZIA, the report would be improved by including the following two long-term open-label extension studies published in peer-reviewed journals evaluating 52 week and 18 month time points.	We did not design this review to look at open label extension trials as there was no comparison group.

Commentator & Affiliation	Section	Comment	Response
Public: UCB	Results, KQ4	Lichtenstein GR, Thomsen OO, Schreiber S et al. Continuous therapy with certolizumab pegol maintains remission of patients with Crohn's disease for up to 18 months. *Clin Gastroenterol Hepatol* 2010;8:600-609.	This study was excluded because it does not have a comparison of interest.
Public: UCB	Results, KQ5	Sandborn WJ, Schreiber S, Hanauer SB, Colombel JF, Bloomfield R, Lichtenstein GR. Reinduction with certolizumab pegol in patients with relapsed Crohn's disease: results from the PRECiSE 4 Study. *Clin Gastroenterol Hepatol* 2010;8:696-702.	This study was excluded because it does not have a comparison of interest.
Public: UCB	Results, KQ6	In the evaluation of patient-reported outcomes during the maintenance of remission on p. 70 and p.71 – Table 15, the following patient-reported outcomes data should be added to the evaluation.	This study is included in our analysis.
Public: UCB	Results, KQ7	Feagan BG, Coteur G, Tan S, Keininger DL, Schreiber S. Clinically meaningful improvement in health-related quality of life in a randomized controlled trial of certolizumab pegol maintenance therapy for Crohn's disease. *Am J Gastroenterol* 2009;104:1976-1983.	This study is included in our analysis.
Public: UCB	Results, KQ8	We have noted the following relevant technical corrections that we recommend are addressed in the final report: (see table in UCB biotech peer comments.	Thank you for these corrections. We are no longer including data from these time points.
Public: UCB	Conclusion	In conclusion, UCB would like to thank AHRQ for the opportunity to comment on this draft report. We appreciate AHRQ's willingness to partner with healthcare stakeholders, including the life sciences industry, to improve our nation's healthcare. As our efforts in this therapeutic area continue, we look forward to further collaboration with the Agency on improving the body of clinical evidence for Crohn's disease and other important therapeutic areas. We are happy to answer any questions you may have on these comments and/or provide additional information as needed.	Thank you for feedback!
Public: Janssen Biotech	General	We suggest that the final comparative effectiveness review be strengthened by summarizing outcome findings relevant from the prescribing information for REMICADE, along with information obtained from the publications provided in the Appendix section of this document. Such enhancements could strengthen the review's utility for the medical community and provide greater clarity of this important topic for policy makers.	We did not make any specific changes to the report in response to this comment.
Public: Janssen Biotech	Abstract	Existing Text: There was high strength of evidence that specific biologics were more effective than placebo at inducing or maintaining remission at some timepoints through 1 year. Adalimumab, certolizumab pegol, and natalizumab were effective at inducing remission. Recommendation: Please add INFLIXIMAB to the following sentence: There was high strength of evidence that specific biologics were more effective than placebo at inducing or maintaining remission at some timepoints through 1 year. Infliximab, adalimumab, certolizumab pegol, and natalizumab were effective at inducing remission.	We have reviewed and revised our strength of evidence to make sure that it is consistent with our methods. The results sentence in our abstract now states, "For adults, infliximab and 6-methylprednisolone were consistently favored over placebo across the induction and maintenance outcomes."

Commentator & Affiliation	Section	Comment	Response
Public: Janssen Biotech	Abstract	Existing Text: For children, the strength of evidence was low or insufficient to support the effectiveness of any medication in inducing or maintaining remission. Recommendation: Please note that INFLIXIMAB is approved for induced and maintained in response and remission in pediatric patients with moderately to severely active UC and CD disease in patients who have failed conventional therapies.	No changes made.
Public: Janssen Biotech	Abstract	Existing Text: Strong evidence exists for the short-term efficacy of adalimumab, certolizumab pegol, infliximab, natalizumab, and budesonide in decreasing disease activity in adults with Crohn's disease, but only weak or insufficient evidence exists on the safety and long-term effectiveness of treatment options. Recommendation: Please remove INFLIXIMAB from the discussion as having weak safety evidence.	We reviewed the strength of evidence grades to make sure that the wording of this text is consistent with the strength of evidence.
Public: Janssen Biotech	Abstract	Existing Text No pediatric study reported on a serious adverse event such as mortality, progressive multifocal leukoencephalopathy, lymphoma, or other cancers. Recommendation: Lymphoma or other cancers have been reported in pediatric patients, by stating in this manner where it only pertains to clinical trial data, may minimize the risk of this important safety topic.	We did not identify observational studies that met our inclusion criteria that reported on these outcomes, so we kept the text as it was written. We did not have room in the abstract to add the additional information recommended by the reviewer.
Public: Janssen Biotech	Executive Summary	Existing Text: Crohn's disease affects between 400,000 and 600,000 North Americans Recommendation: Please update the incidence numbers to: It is estimated that 1.4 million Americans suffer from Crohn's disease or ulcerative colitis Justification: Information according to the Centers for Disease Control and Prevention: http://www.cdc.gov/ibd/	Thank you very much for this suggestion. We decided not to use this reference because it provides prevalence data for IBD, not just Crohn's disease.
Public: Janssen Biotech	Executive Summary, Table A	Existing Test: Infliximab half-life listed as: 9.8 days. Recommendation: please update to: 7.7 to 9.5 days Justification: per the PI for Infliximab.	This has been updated in Tables A and 1.

Commentator & Affiliation	Section	Comment	Response
Public: Janssen Biotech	Executive Summary	Existing Text: We did not include RCTs that examined only the same medication administered at different time points or at different dosages. Recommendation: This is a major flaw in the selection process as dose ranging studies, pharmacokinetic studies, many safety studies invested the agents at various doses to determine safety parameters.	For KQ1 and KQ2 (on induction and maintenance of remission, which are effectiveness outcomes), we included only RCTs because less rigorous study designs would not add much to our ability to estimate the effectiveness of the medications. We did not include RCTs that examined only the same medication administered at different dosages because such studies did not directly address our key questions. However, we did include observational studies so that we could address KQ3 on safety, recognizing that RCTs typically are not designed to assess all important safety related outcomes. We decided not to add more detail to this part of the Executive Summary because it is already quite long.
Public: Janssen Biotech	Executive Summary	Existing Text: We found high strength of evidence for the clinical effectiveness of particular medications at specific time points using a disease activity scale. Statistical significance was not required for an effect to be considered clinically significant, although comparisons with high strength of evidence tended to be statistically significant. For induction of remission, the strength of evidence was high that natalizumab was more effective than placebo at 2-4 weeks, adalimumab was more effective than placebo at 2-4 weeks, and certolizumab pegol was more effective than placebo at 22-30 weeks. For maintenance of remission, the strength of evidence was high that, among those who achieved response or remission during an open-label run-in period (using the study drug), infliximab and adalimumab were more effective than placebo at maintaining remission at 22-30 and 48-54 weeks, and adalimumab and certolizumab pegol was more effective than placebo at 15-19 weeks. Recommendation: Please add in: For induction of remission, infliximab was more effective than placebo at 2-8 weeks.	We re-assessed the strength of evidence focusing only on selected time points and remission. We now state, “For induction of remission, as measured by disease activity, the SOE was high that both natalizumab and TNF-alpha inhibitors (infliximab, adalimumab, and certolizumab pegol) as a class were more effective than placebo for the induction of early remission (weeks 2-4).”

Commentator & Affiliation	Section	Comment	Response
Public: Janssen Biotech	Executive Summary	Existing Text: The applicability to newly diagnosed patients was limited because many of the trials included patients with at least 10 years of Crohn's disease prior to randomization who had previously used medications. Recommendation: Please include the following study: Colombel JF, Sandborn WJ, Reinisch W, et al. Infliximab, azathioprine, or combination therapy for Crohn's disease. N Engl J Med. 2010;362:1383–95.	We included the Colombel study in our review.
Public: Janssen Biotech	Executive Summary	Existing Text: Pediatric results We did not find any RCT of biologics in children that met the criteria for inclusion in this review. An RCT of maintenance therapy with on-demand infliximab in pediatric patients was not included because there was no comparison with another medication, only with the timing of infliximab administration. Recommendation: Please consider adding in and evaluating two studies to this review regarding infliximab in pediatric patients with moderately to severely active CD and UC. Infliximab does have the following FDA indications in pediatric patients: INFLIXIMAB is indicated for reducing signs and symptoms and inducing and maintaining clinical remission in pediatric patients 6 years of age and older with moderately to severely active Crohn's disease who have had an inadequate response to conventional therapy. INFLIXIMAB is indicated for reducing signs and symptoms and inducing and maintaining clinical remission in pediatric patients 6 years of age and older with moderately to severely active ulcerative colitis who have had an inadequate response to conventional therapy.	These sentences have been removed from the Pediatric Results section of the Executive Summary.
Public: Janssen Biotech	Executive Summary	Existing Text: Table C: Summary of the strength of evidence on the comparative effectiveness of pharmacologic therapies for the management of Crohn's disease in terms of inducing a remission as measured by a disease activity index at various time points. à Weeks 8-12 (induction) Recommendation: Please note the US Food and Drug Administration considered the induction period for infliximab to be 0, 2 and 6 weeks per the PI for infliximab.	For purposes of this study, early remission is defined as 2-4 weeks and later remission is defined at 12-16 weeks. We did not consider 6-week data. This is described in the Data Analysis and Synthesis section of the Methods.
Public: Janssen Biotech	Executive Summary	Existing Text: Table C: Summary of the strength of evidence on the comparative effectiveness of pharmacologic therapies for the management of Crohn's disease in terms of inducing a remission as measured by a disease activity index at various time points. à Weeks 2-4 (immediate response) = Insufficient. Recommendation: Please consider changing the SOE to Mod or Good based on data below. Colombel et al conducted a multicenter, randomized, double-blind, controlled trial comparing INFLIXIMAB monotherapy versus INFLIXIMAB plus azathioprine (AZA) versus AZA alone for the treatment of moderate to severe Crohn's disease (Crohn's Disease Activity Index [CDAI] score ≥220 and ≤450) in patients naïve to both immunomodulators and biologic therapy.	We relied on our defined approach to grade the strength of evidence. As it pertains to the Colombel et al study, first time points are given at week 6. This is not a very clinically useful time point with regards to comparison with azathioprine which takes 12 weeks or longer to work.
Public: Janssen Biotech	Executive Summary	Recommendation: Infliximab is high SOE for the management of Crohn's disease in terms of maintaining remission as measured by a disease activity index at Weeks 15-19. Please adjust this section of the table to reflect this accurately.	We are no longer looking at this time point for KQ2. The first time point will be 1 year.

Commentator & Affiliation	Section	Comment	Response
Public: Janssen Biotech	Executive Summary	Recommendation: Please change the SOE of Infliximab to High for the following endpoints: Induction of remission: Mucosal Healing; Fistula Response Maintenance of remission: Mucosal Healing, Hospitalizations, Fistula response, Patient Reported Outcomes	The epidemiologist and clinical expert re-evaluated the risk of bias and re-graded all strength of evidence for the biologic trials for the induction and maintenance of remission key questions. We did not re-evaluate the risk of bias for other medications or for the safety outcomes. We did re-consider the SOE grading for all comparisons for the induction and maintenance of remission key questions.
Public: Janssen Biotech	Executive Summary	Infliximab vs. azathioprine comparison. Recommendation: Please consider changing the Strength of evidence to High for the column above. Justification: In the study of biologic and immunomodulator naïve patients in crohn's disease) (SONIC) (Colombel et al, 2010; Supplement to: Colombel et al, 2010) - primary efficacy endpoint was the proportion of patients in corticosteroid-free clinical remission (CDAI <150) at week 26. Corticosteroid free clinical remission was defined as clinical remission in patients who had not received budesonide ≥6 mg/day or systemic corticosteroids for at least 3 weeks. Additional secondary endpoints included mucosal healing at week 26 in patients with mucosal ulcerations present at baseline as well as the proportion of patients in corticosteroid-free clinical remission at other time points including week 50. Also evaluated were the rate of clinical remission, response-70, response-100, and the Inflammatory Bowel Disease Questionnaire (IBDQ) score. Please see the selected results from SONIC below.	Thank you for this suggestion. We have graded the strength of evidence for these comparisons and outcomes using the Methods described in the Rating the Body of Evidence section. The Colombel 2010 study was included in our assessment of the evidence grade.
Public: Janssen Biotech	Executive Summary	Existing Text: The major limitations in our review were: (a) the inability to perform meta-analyses of all comparisons, (b) potential measurement error, and (c) no established standard for a clinically meaningful difference in remission or safety related outcomes. Recommendation: Please consider adding to the limitations statement. Ranking of clinical trials used as part of the US Food & Drug Administration approval process something other than high strength of evidence.	We did not make any specific changes in response to this comment. We would like for our review of the evidence to be considered independent of the FDA's process.
Public: Janssen Biotech	Results, KQ1	Existing Text: The strength of evidence was moderate to demonstrate that one dose of infliximab was more effective than placebo in inducing a response, but not remission, at weeks 2 and 12 (absolute RD across time points, 10% to 33%; placebo rate, 4% to 8%). Recommendation: Please consider changing the strength of evidence to High	We have graded the strength of evidence for these comparisons and outcomes using the Methods described in the Rating the Body of Evidence section.

Commentator & Affiliation	Section	Comment	Response
Public: Janssen Biotech	Results, KQ1	Existing Text: The strength of evidence was moderate to demonstrate that infliximab (5 mg/kg induction and maintenance) was more effective than azathioprine (2.5 mg/kg/day) in inducing a steroid-free remission at weeks 10 and 26 (absolute RD across time points, 13% to 16%; azathioprine rate, 24% to 32%). The strength of evidence was moderate to demonstrate that a combination of infliximab (5 mg/kg induction and maintenance) and azathioprine (2.5 mg/kg/day) was more effective than infliximab (5 mg/kg induction and maintenance) alone in inducing a steroid-free remission at weeks 10 and 26 (absolute RD across time points, 10% to 12%; infliximab alone rate, 37% to 48%). The strength of evidence was moderate to demonstrate that a combination of infliximab (5 mg/kg induction and maintenance) and a thiopurine (azathioprine or 6- mercaptopurine) was more effective than a thiopurine alone in inducing a steroid-free remission at weeks 10 to 12 and weeks 24 to 26 (absolute RD across time points, 12% to 30%; azathioprine rate, 24% to 48%). Recommendation & Justification: Please consider adjusting the SOE.	This was rated as moderate strength of evidence because it excluded patients with below normal thiopurine methyltransferase thus making it less likely that azathioprine would be successful.
Public: Janssen Biotech	Results, KQ1	Existing Text: The strength of evidence was moderate to demonstrate that infliximab was more effective than placebo in achieving mucosal healing at week 4 (absolute between-group difference in the change from baseline in Crohn's disease endoscopic index of severity [CDEIS] score, 7.7; placebo difference, 0.9). The strength of evidence was moderate to demonstrate that infliximab (5 mg/kg induction and maintenance) and placebo was more effective than azathioprine (2.5 mg/kg daily) in achieving mucosal healing at week 26 (absolute RD in percentage of patients who achieved absence of mucosal ulcers, 13%; azathioprine and placebo rate, 17%). The strength of evidence was moderate to demonstrate that a combination of infliximab (5 mg/kg induction and maintenance) and azathioprine (2.5 mg/kg daily) was more effective than infliximab (5 mg/kg induction and maintenance) and placebo in achieving mucosal healing at week 26 (absolute RD in percentage of patients who achieved absence of mucosal ulcers, 14%; infliximab and placebo rate, 30%). The strength of evidence was moderate to demonstrate that a combination of infliximab (5 mg/kg induction and maintenance) and azathioprine (2.5 mg/kg daily) was more effective than azathioprine (2.5 mg/kg daily) and placebo in achieving mucosal healing at week 26 (absolute RD in percentage of patients who achieved absence of mucosal ulcers, 27%; azathioprine and placebo rate, 17%). Recommendation & Justification: Please consider adjusting the SOE.	We have graded the strength of evidence for these comparisons and outcomes using the Methods described in the Rating the Body of Evidence section.
Public: Janssen Biotech	Results, KQ1	Existing Text: The strength of evidence was moderate to demonstrate that infliximab was more effective than placebo in healing fistulas at week 6 in patients with actively draining fistulas (absolute RD in fistula closure, 25% to 42%; placebo rate, 13%). Recommendation & Justification: Please consider adjusting the SOE.	We have graded the strength of evidence for these comparisons and outcomes using the Methods described in the Rating the Body of Evidence section.

Commentator & Affiliation	Section	Comment	Response
Public: Janssen Biotech	Results, KQ1	Recommendation & Justification: Please consider adjusting the remission rate numbers to those from the current prescribing information for Infliximab – include Accent 1 and Accent 2 trials. Please use the published supplementary index for Sonic for the Response Rate (%) (100-pt CDAI drop) and Response Rate (%) (70-pt CDAI drop)	We have included ACCENT I and II in our report. We are no longer looking at response rates.
Public: Janssen Biotech	Results, KQ1	Existing Text: Biologics versus placebo. One study with 30 participants (a subgroup of patients from Targan et al.)77 compared infliximab with placebo for this outcome.118 Crohn's disease endoscopic index of severity (CDEIS) was significantly improved in the infliximab group (all doses, from 5 mg/kg to 20 mg/kg) at 4 weeks compared to baseline (mean change of 7.7; P < 0.001). The mean change in CDEIS for placebo was 0.9 (P = NS). Notably, the baseline CDEIS was higher in the infliximab group (13.0) compared to placebo (8.4). Recommendation & Justification: Please consider incorporating information from the ACCENT 1 and Sonic Study regarding Mucosal Healing.	We have graded the strength of evidence for these comparisons and outcomes using the Methods described in the Rating the Body of Evidence section. These studies are included in the analysis.
Public: Janssen Biotech	Results, KQ2	Existing Text: The strength of evidence was low that infliximab was more effective than placebo at week 16 (absolute RD, 30%; placebo rate, 30%). Recommendation & Justification: Please consider deleting bullet point in the above text and pertaining reference to this data for Week 16 efficacy – as it is incomplete information to have an efficacy marker that is applicable only to one biologic agent and not all.	We have eliminated this time point.
Public: Janssen Biotech	Results, KQ2	Existing Text: The strength of evidence was low to demonstrate that there was no difference between the combination of infliximab (5 mg/kg every 8 weeks) and azathioprine (2-2.5 mg/kg daily) and infliximab (5 mg/kg every 8 weeks) with intravenous hydrocortisone pretreatment in maintaining remission at 6, 12, and 24 months among patients with steroid-dependent disease (range in absolute RD across time points, -9% to -5%; infliximab and hydrocortisone rate range, 77% to 88%). Recommendation & Justification: Please consider revising bullet point in the above text.	This comment does not provide a strong enough rationale for changing our recommendation.
Public: Janssen Biotech	Results, KQ2	Existing Text: The strength of evidence was moderate to demonstrate that infliximab was superior to placebo in closing fistulas at week 40 in adults with Crohn's disease who had achieved an initial fistula response to open-label drug (absolute RD, 17%; placebo rate, 19%). Recommendation & Justification: Please consider revising the SOE to HIGH	We have graded the strength of evidence for these comparisons and outcomes using the Methods described in the Rating the Body of Evidence section.
Public: Janssen Biotech	Results, KQ2	Existing Text: Mucosal Healing Recommendation & Justification: Please consider including the Colembel SONIC study	Data from the Colombel study are included under KQ1.
Public: Janssen Biotech	Results, KQ2	Existing Text: Reduction of Steroids Recommendation & Justification: Please consider including the Colembel SONIC study	Data from the Colombel study are included under KQ1.
Public: Janssen Biotech	Results, KQ3	Please include the Warnings and Precautions section of each biologic agent along with the black box warnings (if applicable). Please see the citations provided in the Appendix of this document.	We added Table 2 to the Introduction, which lists the black box warnings for each of the included therapies.

Commentator & Affiliation	Section	Comment	Response
Public: Janssen Biotech	Results, KQ4	Hyams et al, 2002. Induction and maintenance infliximab therapy for the treatment of moderate-to-severe Crohn's disease in children. The primary objective of this trial was to evaluate the efficacy of a 3-dose induction regimen of REMICADE in reducing signs and symptoms in pediatric patients with also evaluated. The secondary objective was to compare maintenance of clinical response and emission with REMICADE 5 mg/kg administered every 8 weeks or every 12 weeks. Hyams et al conducted a tandomized, open label, controlled trial to determine the safety and efficacy of REMICADE in 112 pediatric patients with active CD. Of the 112 patients enrolled in the study, 103 patients achieved clinical response to induction therapy (REMICADE 5 mg/kg at 0, 2, 6 weeks) and were subsequently randomized (1:1) to REMICADE 5 mg/kg every 8 weeks (n=52) or every 12 weeks (n=51) through week 46. Patients who lost their-point increase in the PCDAI (from baseline to week 10) occurring ≥7 days apart at 2 consecutive follow-up visits, or a PCDAI > 30 week maintenance therapy were eligible to crossover to 10 mg/kg every 8 weeks in the 5 mg/kg every 12 week maintenance therapy group, patients who lost response ≤8 weeks following their last infusion received subsequent dosing with 10 mg/kg every 8 weeks. Those patients who lost response >8 weeks but ≤12 weeks received 5 mg/kg every 8 weeks. Baseline disease characteristics and demographics were similar between groups. Additionally, at baseline the mean age was 13.3 ± 2.5 years (range, 6-17) and the mean PCDAI score was 41.2 ± 8.3. The overall discontinuation rate observed in REACH was 21.4% (24/112). Of the 24 subjects, 9 discontinued prior to or at week 10 and 15 after week 10. In the 15 subjects who discontinued after week 10, a higher discontinuation rate was observed in the every 12-week maintenance group than in the every 8-week maintenance group.	This study was excluded because it does not have a comparison of interest.
Public: Janssen Biotech	Results, KQ4	Clinical response at week 10, the primary endpoint, was achieved in 88% of the patients. Additional response and remission rates are summarized in ERROR Ref. The data shown for maintenance of response and remission at week 54 is based on intention-to-treat analysis with patients who crossed over to a higher or more frequent REMICADE dose counted as treatment failures. In the combined every 8 and 12 weeks maintenance regimens, at weeks 10, 30, and 54, the decrease from baseline in average daily corticosteroid use was significant ($p<0.001$ for all timepoints). Additionally, patients with a ≥1 year delay in bone age (mean z-score at baseline = -1.5) were evaluated in this study. In these patients, the mean change from baseline in height at weeks 30 and 54 was 0.3 cm ($p<0.001$) and 0.5 cm ($p<0.001$), in the combined REMICADE groups, respectively. Furthermore, quality of life was assessed in 76 patients from North America (aged 10-17, median IMPACT III score of 90). Baseline improvements in IMPACT III increased an average of 23.9 at week 10 ($p<0.001$). Similarly, improvements were noted at week 30 and 54 ($p<0.001$ for both time points).	This study was excluded because it does not have a comparison of interest.

Commentator & Affiliation	Section	Comment	Response
Public: Janssen Biotech	Results, KQ4	In REACH, the proportion of patients experiencing adverse events and serious adverse events was similar between the two maintenance regimens, including upper respiratory tract infections and anemia. Infections were reported more frequently for subjects who received every 8-week as opposed to every 12 week infusions (74% and 38%, respectively), while serious infections were reported for 3 patients in the every 8 week and 4 patients in the every 12 week maintenance treatment group. Pneumonia was reported for 3 patients, (2 in the every 8 week and 1 in the every 12 week maintenance treatment groups). Herpes zoster was reported for 2 patients in the every 8-week maintenance treatment group. Eighteen percent of randomized patients experienced one or more infusion reactions, with no notable difference between treatment groups. Of the 112 patients in REACH, 2 patients had possible anaphylactic reactions. Antibodies to REMICADE developed in 3% of ≥150 U/L) were seen in 6% of patients receiving REMICADE every 12 weeks (observed during the induction phase), levels did return to normal during maintenance therapy. No marked elevations in AST were reported. One patient receiving REMICADE every 8 weeks developed an elevated bilirubin level, which returned to normal despite continued REMICADE therapy. No deaths, malignancies, demyelinating disorders, optic neuritis, seizures, or new cases of autoimmune disease were reported.	This study was excluded because it does not have a comparison of interest.
Public: Abbott	Overall	Consider reevaluating study inclusion criteria or another method in assessing pediatric Crohn's Disease patients due to the many challenges in studying this unique population. Challenges include study design, sample size and duration of study limited due to safety and ethical concerns associated with studying this population.	Thank you very much for your feedback. We did not make the recommended change because we thought it was important to be consistent in how we defined the study inclusion criteria for adult and pediatric patients.
Public: Abbott	Overall	Consider specifying type of CD approval in Table A (List of medications used for the treatment of CD). Table A only indicates approval for CD, but does not distinguish between induction and/or maintenance of remission and/or major clinical response.	Thank you for this suggestion, but we decided not to add this additional detail to the report.
Public: Abbott	Overall	Consider reevaluating the power of clinically meaningful differences (using a threshold of 10%) of remission and safety outcomes. Because there is no standard for assessing clinically meaningful differences, true clinically meaningful differences can not be determined without meeting statistical significance, as well. Otherwise, the results could be due to chance.	As a team, we decided on 10% as clinically meaningful prior to evaluating the evidence. No changes were made.
Public: Abbott	Overall	Onset of action varies from 4 to 8 weeks for the biologics versus 3 to 12 months for immunosuppressants. Please consider how this difference will factor into induction of remission for comparisons. A two week time point for induction of remission may be too conservative considering variation of onset of action.	Agreed. Thus, with the early time points for immunosuppressants we clearly point out that there is an issue of applicability as pertains to their duration of onset
Public: Abbott	Overall	Consider adding citations to all tables and figures within the Draft Report, as verifying data is challenging without references.	Citations are in the tables.

Commentator & Affiliation	Section	Comment	Response
Public: Abbott	Executive Summary	Please clarify the statement on induction of remission in adults for adalimumab. The statement reads: "For induction of remission, the strength of evidence was high that infliximab was more effective than placebo at 4 weeks and that neither medication was favored comparing natalizumab to placebo at 12 weeks, adalimumab to placebo at 4 weeks, and certolizumab pegol to placebo at 12 and 26 weeks." Adalimumab is more effective than placebo for induction of remission. Please see references.	We have revised our conclusions for KQ1 and no longer specifically mention adalimumab in the text. The results are presented in Table C. We have also reviewed our evidence grading to make sure that it is consistent with our Methods.
Public: Abbott	Results, KQ1	In Figure 2, please make mathematical correction or identify unaccounted for abstracts as 4868 abstracts minus 3857 equals 1011, not 997 (14 unaccounted).	This has been updated.
Public: Abbott	Results, KQ1	In the mucosal healing section, consider including the EXTEND trial.	Thank you for this suggestion. We excluded the EXTEND trial because we excluded studies that were only published as abstracts.
Public: Abbott	Results, KQ1	In the hospitalization section, consider including the CHARM hospitalization data, which showed the significant benefit of adalimumab over placebo.	The data from the Feagan 2008 regarding hospitalizations is included under KQ2.
Public: Abbott	Results, KQ1	In the reduction of steroids section, consider including steroid-free remission data from the CHARM study published by Kamm, et al., which showed the significant benefit of adalimumab over placebo.	The Kamm 2011 study was included in our updated search.
Public: Abbott	Results, KQ1	In the fistula section, consider including fistula healing data from CHARM trial published by Colombel, et al., which showed the significant benefit of adalimumab over placebo.	We only included fistula maintenance studies in which patients initially had fistula healing, and then were randomized to drug versus placebo. Only ACCENT II had this study design; CHARM and PRECISE II did not. Additionally, for the reference included in this comment, this summarized the subgroup analysis from CHARM and additionally presented data on open-label extension treatment with adalimumab after week 56. As there was no comparison group in this extension study, it did not meet our criterion for inclusion.
Public: Abbott	Results, KQ1	Consider correction of statement of PROs for adalimumab, where adalimumab treatment showed statistical differences over placebo, yet statement suggests no difference.	As described in our methods, the difference between treatments is not clinically meaningful.

Commentator & Affiliation	Section	Comment	Response
Public: Abbott	Results, KQ1	In the fistula response section, consider adding underlined insertion to the following statement, "There was no clinical or statistical difference in fistula response (75%, 20%, 8%, and 33% at 40 mg/20 mg, 80 mg/40 mg, 160 mg/80 mg doses, and placebo, respectively) or complete fistula closure (75%, 0%, 0%, and 17%, respectively) for adalimumab versus placebo at week 4".	Thank you. This has been added.
Public: Abbott	Results, KQ2	In the Disease Activity Index for TNF-Alpha Inhibitors, consider clarifying this statement as it currently suggests that all three products have high evidence for remission out to 28 and 52 weeks. Per the subsequent statements and Table D, only adalimumab and infliximab have suffficient data at both of these later time points, not certolizumab.	We clarified that certolizumab pegol does not have 1 year data.
Public: Abbott	Results, KQ2	Consider including mucosal healing data for adaliumumab.	Thank you for this suggestion. We are not including data from abstracts in our review.
Public: Abbott	Results, KQ2	In fistula response section, fistula response data for adalimumab in Draft Report deemed insufficient. Consider data on fistula response from CHARM, which showed the significant benefit of adalimumab over placebo.	We only included fistula maintenance studies in which patients initially had fistula healing, and then were randomized to drug versus placebo. Only ACCENT II had this study design; CHARM and PRECISE II did not. Additionally, for the reference included in this comment, this summarized the subgroup analysis from CHARM and additionally presented data on open-label extension treatment with adalimumab after week 56. As there was no comparison group in this extension study, it did not meet our criterion for inclusion.
Public: Abbott	Results, KQ2	Consider reassessment of the PRO statement for adalimumab, as there is statistically significant benefit of adalimumab over placebo in long-term improvements of patient-reported outcomes.	This study is included in our analysis.

Commentator & Affiliation	Section	Comment	Response
Public: Abbott	Results, KQ2	In Table 15, consider revising to change evidence grade for fistula data with adaliumumab. Please see reference.	We only included fistula maintenance studies in which patients initially had fistula healing, and then were randomized to drug versus placebo. Only ACCENT II had this study design; CHARM and PRECISE II did not. Additionally, for the reference included in this comment, this summarized the subgroup analysis from CHARM and additionally presented data on open-label extension treatment with adalimumab after week 56. As there was no comparison group in this extension study, it did not meet our criterion for inclusion.
Public: Abbott	Results, KQ2	The Draft Report states on page 134 that "Although both medications have a unique make-up, we believe that the mechanism of action for all three is similar." However, the pooled analysis in Figure 7 reports only adalimumab and infliximab. This statement is contradictory to the data being evaluated and is confusing. Please reconsider revising the statement to refer to only the two products that are being evaluated, adalimumab and infliximab.	We clearly state we are evaluating all 3 anti-TNFs.
Public: Abbott	Results, KQ2	The Draft Report only analyzes data on adalimumab and infliximab (not all Anti-TNFs) for remission at 52 weeks. Please reconsider revising the statements within this section to refer only to the two products that are being evaluated, adalimumab and infliximab.	We refer only to adalimumab and infliximab for maintenance of remission at 52 weeks in the KQ2 results of the main report.
Public: Abbott	Results, KQ2	In surgeries section, adalimumab vs placebo, 854 patients received open-label induction therapy with adalimumab in CHARM, not 778. Before randomization, 76 patients withdrew, where remaining 778 patients were stratified. Please make correction.	This figure has been removed from the text.
Public: Abbott	Results, KQ2	In fistula response section for adalimumab, consider adding "of patients who had complete closure at week 26, 100% maintained complete closure at week 52".	Although this may be true, we only studied patients at the time of randomization; thus, this would not be an endpoint we would look at.
Public: Abbott	Results, KQ2	With regard to IBDQ scores for adalimumab, the draft report states nonstatistically significant change in scores from week 4 to week 52 (after the open-label run-in). Consider reporting also on the change in IBDQ scores from during the induction phase as well which was statistically and clinically significant.	For KQ2, we are focusing on the results for patients that have inactive disease at the point of randomization.

Commentator & Affiliation	Section	Comment	Response
Public: Abbott	Results, KQ3	Consider distinguishing between injection site reactions for drugs given subcutaneously (adalimumab, certolizumab, ustekinimab) versus infusion reactions for drugs that are give intravenously (infliximab, natalizumab). Infusion reactions can be severe or life threatening, whereas injection site reactions are generally transient and not considered severe.	We agree that there is a spectrum of severity for infusion and injection site reactions. However, the body of literature does not distinguish severity of reactions well. We have tried to be specific as to severity when possible.
Public: Abbott	Results, KQ3	Title of Table 26 states "Trials that did not report on mortality", which is different from outcome in column 3 "Trials reporting no deaths". Please clarify which of these two outcomes are being included in the table for consistency.	We have changed the title of the table to be, "Summary of randomized controlled trials that reported no deaths when comparing the effectiveness of a biologic alone or in combination with placebo or another treatment in patients with Crohn's disease."
Public: Abbott	Results, KQ3	In Table 26, comparison group, column 2, states placebo was the comparison for adalimumab study. To be accurate, the studies used placebo plus any background immunosupressives that patients were receiving upon study entry. It would be generally not be possible to conduct a study that removes all treatment except for placebo as a control group in treating Crohn's disease.	This is a study characteristic common to many studies and described in the study characteristics. We could not provide the additional information about background immunosuppressives because of the limitations in how this was reported in the literature.
Public: Abbott	Results, KQ3	In Table 25, please include adalimumab randomized placebo-controlled trials, CHARM, CLASSIC, and GAIN. Please see references.	These are reported in Table 26.
Public: Abbott	Results, KQ3	Consider clarifying in the adalimuamb versus plavebo section that adalimumab trials did not compare strictly to placebo alone but compared to placebo plus any background immunosupressives patients were on at study entry.	This is true of almost all RCTs included and is considered in study characteristics and discussion.
Public: Abbott	Results, KQ3	In Hepatosplenic T-Cell Lymphoma section, important to recognize some of the limitations of the AERS database with regards to the search strategy for Hepatosplenic T-Cell Lymphoma. Consider adding text relative to limitations of AERS database.	No changes made. This suggestion is beyond the scope of this report, but we are working on a separate report that examines in more detail data from the AERS database.
Public: Abbott	Results, KQ3	In Table 33, Column header of "Mean Number of Infusions" is misleading since not all the medications mentioned are actually infused. Mean number of doses may be more accurate. Consider making that revision.	We changed this to injections or infusions.

Commentator & Affiliation	Section	Comment	Response
Public: Abbott	Results, KQ3	In Figure 21, for the Hanauer 2006 adalimumab study, N of TNF = 225, not 223; For the Colombel 2007 & Sandborn 2007adalimumab studies, it should read 56 weeks, not 52 weeks; Consider revising values. Consider citing data to verify data contained in reference.	We have corrected the number of patients included in the TNF arm for the Hanauer 2006 adalimumab study. For all studies, we are only presenting the followup since randomization. For the Colombel 2007 and Sandborn 2007 studies, this is 52 weeks (56-4 week induction=52).
Public: Abbott	Results, KQ3	In Figure 23, for Colombel 2009 adalimumab study, comparison states placebo which seems misleading as it was against induction only and then reinitiation; should state 56 weeks and not 52 weeks. Consider citing data to verify data contained in reference. Several studies were publised by Colombel in 2009.	Although this may be true, we needed to simplify the comparisons for presentation in the figure. For all studies, we are only presenting the followup since randomization. For the Colombel 2007 and Sandborn 2007 studies, this is 52 weeks (56-4 week induction=52).
Public: Abbott	Results, KQ3	In Figure 24, please define the abbreviation for TEI. Not listed anywhere in the report or the glossary of terms.	The abbreviation has been defined in the footnote.
Public: Abbott	Results, KQ3	In Figure 24, Colombel 2009 adalimumab study should read 56 weeks, not 52 weeks. Please make this correction.	For all studies, we are only presenting the followup since randomization. For the Colombel 2007 and Sandborn 2007 studies, this is 52 weeks (56-4 week induction=52).
Public: Abbott	Results, KQ3	In the adalimumab versus placebo section, potentially misleading characterization of placebo patients in ref 129 (Columbel 2007) and ref 135 (Colombel 2009) as far as serious infections (4% in adalimumab and 5% in placebo. The placebo in this trial was all patients receiving two 80mg adalimumab doses over a 2 week period and then went through a blinded switch to placebo).	All patients in CHARM received 80 mg week 0, followed by 40 mg week 2, then randomized to continue drug or get placebo. We agree that this is not pure drug vs. placebo comparison, but more of a comparison of episodic versus maintenance strategy.
Public: Abbott	Results, KQ3	Unlike other treatment classes, there is no figure for Peto Odds Ratio. A graphic of risk for serious infections or opportunistic infections for steroids would be informative in understanding and communicating the magnitude of risk for this class of drugs as the the report did for other drug classes.	None of the studies evaluating corticosteroids reported on opportunistic infections, so we can't make a figure displaying that data.

Commentator & Affiliation	Section	Comment	Response
Public: Abbott	Results, KQ3	In Tuberculosis section, please consider correcting the statement, "The FDA has mandated a warning on infliximab and adalimumab regarding the risk of tuberculosis". The report fails to state that ALL TNF-alpha inhibitors, including certolizumab, are mandated by FDA to carry class level box warning regarding risk for tuberculosis. Please update the report to reflect that all 3 TNF-alpha inhibitors have a black box warning on the risk of tuberculosis.	We added a citation for certolizumab.
Public: Abbott	Results, KQ3	In Table 44, column 5 the following adalimumab studies are described incorrectly. These studies only exclude patients who had a history of active tuberculosis (Tb) infection. Patients who tested PPD positive, but did not have a history active Tb were allowed into these adalimumab trials so long as they took Tb prophylaxis.	We changed this to tuberculosis testing instead of "exclusion of tuberculosis."
Public: Abbott	Results, KQ3	In Table 44, column 5 the following adalimumab studies are described incorrectly. These studies did not exclude patients who were screened positive for tuberculosis (Tb) with a PPD skin test. Patients who tested PPD positive, but did not have active Tb were allowed into these adalimumab trials so long as they took Tb prophylaxis and did not have active Tb disease. These studies include reference numbers 79 (Sandborn 2007) and 135 (Colombel 2009).	We have changed the column label.
Public: Abbott	Results, KQ3	In Table 44, column 5 the following adalimumab studies deserve clarification. The adalimumab studies (ref # 78, 79, 130, 135) applied the inclusion/exclusion criteria equally between the treatment and control groups within each study. Patients with a history of exposure to tuberculosis as demonstrated by a positive PPD skin test could be randomized to be in the adalimumab group, and they as likely could have been randomized to assignment into the placebo group so long as they were willing to take Tb prophylaxis medication. In other words, patients were not selected out of either arm of study based on PPD positivity.	We changed this to tuberculosis testing instead of "exclusion of tuberculosis."
Public: Abbott	Results, KQ3	In Table 45, Column 2, row 3 incorrectly states this study as being unblinded. The CHARM trial was a double blind study.	We have made this change. After 12 weeks, patients could change therapy.
Public: Abbott	Results, KQ3	In Table 46, Column 2, row 3 incorrectly states this study as being 52 weeks duration. The CHARM trial was a 56 week study.	For all studies, we are only presenting the followup since randomization. For this study, this is 52 weeks (56-4 week induction=52).
Public: Abbott	Results, KQ3	In Figure 29, please verify that ORs reported for all products include true injections site reactions because various products report on only injection site pain. Injection site reactions is a broader term that encompasses injection site pain so comparing them inherently shows a significant difference. This difference merits at least a mention in the table footnote.	We added footnotes to the figure which has definitions on how each study defined injection site reactions.

Commentator & Affiliation	Section	Comment	Response
Public: Abbott	Results, KQ3	Two studies were identified to answer KQ4. They were well represented and characterized acurately. A review of 5-ASA in August 2010 identified 11 studies in post-operative care prevention. Consider reviewing more trials for this Question to give more effectiveness comparisons. Please include references listed.	The Ford study was excluded because there was no original data. The Caprilli 1994 and Hanauer 2004 studies were excluded because they did not report on patient-reported outcomes.
Public: Abbott	Results, KQ3	Consider including the trials for sulfasalazine and 6-MP in the review.	The Ewe 1989 and Hanauer 2008 studies were excluded because they did not report on patient-reported outcomes.
Public: Abbott	Results, KQ4	The clinical significance of the IBDQ and VAS scales are questioned for their applicability. The IBDQ is a validated scale and offers value if used regularly in clinical practice.	We did not make any edits based on this comment.
Public: Abbott	Results, Pediatrics	Consider including pediatric CD induction of remission, maintenance of remission, and safety data studying adalimumab. Please see references listed.	We excluded the Viola 2009, Rosh 2009, and Wyneski 2008 studies because there was no comparison group. We excluded the Rosh 2009 Abstract 1458 study because it is an abstract. We excluded the Hadziselimovic study because it was a case series.
Public: Abbott	Discussion	Consider making mention that of the available options for moderate to severe disease, anti-TNFs have a very positive risk benefit profile.	We did not make any specific changes to the discussion in response to this suggested editorial comment.
Public: Abbott	Discussion	Please note that the data from CHARM support efficacy of adalimumab after 4 weeks. It is not consistent with the adalimumab clinical trial to only highlight 2-4 weeks and 15-19 weeks.	We did not make any specific changes to the discussion in response to this comment. We decided to keep the focus on the time points that we identified in Table B of the Executive Summary.
Public: Abbott	Appendix C	No subjects with CD listed as reason for exclusion. Study included 21 CD pediatric patients. Consider including trial for Pediatic KQs 1 and 2.	This study was excluded because it does not have a comparison of interest.
Public: Abbott	Appendix C	No subjects with CD listed as reason for exclusion. Study included 18 pediatric CD patients. Consider including trial for Pediatric KQs 1 and 2.	We were unable to retrieve this article. It's unclear if the citation is correct.

Commentator & Affiliation	Section	Comment	Response
Public: Abbott	Appendix C	No subjects with CD listed as reason for exclusion. 37 Pediatric CD patients included in trial. Consider including trial for Pediatric KQ 1 and 2.	This study was excluded because it does not have a comparison of interest.
Public: Abbott	Appendix C	No subjects with CD listed as reason for exclusion. All patients in global safety summary were Crohn's Disease patients. Consider including in Report.	Many of the studies included in this manuscript already meet our inclusion criteria. Therefore, including this manuscript would double-count many of the adverse events.
Public: Abbott	Appendix C	No subjects with CD listed as reason for exclusion. Only Pediatric CD patients included in trial. Consider including trial for Pediatric KQ 1 and 2.	This study was excluded because it was a case report. We assume you missed this information, but it is listed as such in Appendix C.
Public: Abbott	Appendix C	Reconsider inclusion of studies based on overall pediatric study methodology comment above. Please see references.	This study was excluded because it was a case report. We assume you missed this information, but it is listed as such in Appendix C.
Public: Abbott	Appendix C	No subjects with CD listed as reason for exclusion. Only Pediatric CD patients included in trial. Consider including trial for Pediatric KQ 1and 2.	This study was excluded because it does not have a comparison group. We assume you missed this information, but it is listed as such in Appendix C.
Public: Abbott	Appendix C	Reconsider inclusion of studies based on overall pediatric study methodology comment above. Please see references.	This study was excluded because it does not report side effects by medication. We assume you missed this information, but it is listed as such in Appendix C.
Public: Abbott	Appendix C	No subjects with CD listed as reason for exclusion. All patients were CD patients. Consider including in report.	This study was excluded because it does not have a comparison group. We assume you missed this information, but it is listed as such in Appendix C.
Public: Abbott	Appendix C	No subjects with CD listed as reason for exclusion. All patients were CD patients. Consider including in report.	This study was excluded because it does not have a comparison group. We assume you missed this information, but it is listed as such in Appendix C.

Commentator & Affiliation	Section	Comment	Response
Public: Abbott	Appendix C	Reconsider inclusion of studies based on overall pediatric study methodology comment above. Please see references.	This study was excluded because it does not have a comparison group. We assume you missed this information, but it is listed as such in Appendix C.
Public: Abbott	Appendix C	The Sandborn WJ et al OLE study with adalimumab after infliximab treatment was excluded due to it being an open label extension. However, safety points in this study may have been useful since it has longer term data. Please consider inclusion of this safety data.	This study was excluded because it does not have a comparison group. We assume you missed this information, but it is listed as such in Appendix C.
Public: Abbott	Appendix C	No subjects with CD listed as reason for exclusion. Only Pediatric CD patients included in trial. Consider including trial for Pediatric KQ 1 and 2.	This study was excluded because it does not have a comparison group. We assume you missed this information, but it is listed as such in Appendix C.
Public: Abbott	Appendix C	Reconsider inclusion of studies based on overall pediatric study methodology comment above. Please see references.	This study was excluded because it does not have a comparison group. We assume you missed this information, but it is listed as such in Appendix C.
Public: Abbott	Appendix C	Reconsider inclusion of studies based on overall pediatric study methodology comment above. Please see references.	This study was excluded because it does not have a comparison group. We assume you missed this information, but it is listed as such in Appendix C.
Public: Abbott	Appendix C	Reconsider inclusion of adalimumab studies. Please see references.	Both of these articles were excluded because they did not have a comparison group.
Public: Jill P. Smith, MD	Executive Summary	There have been 2 landmark papers published regarding novel studies using low dose naltrexone, an opioid receptor antagonist, in adults with active Crohn's disease: Am J Gastroenterology: 102:1-9, 2007 and Dig Dis & Sci. 56:2088-2089, 2011. These studies showed > 85% improved clinical activity of Crohn's disease and the latter study also demonstrated mucosal healing. Naltrexone is taken once a day by mouth and has minimal side effects. The FDA has also recently granted naltrexone Orphan Drug status for its use in pediatrics with Crohn's disease.	Naltrexone was not a medication of interest.
Public: Jill P. Smith, MD	Introduction	Naltrexone is an opioid receptor antagonist that was shown in an animal model of inflammatory bowel disease to improve GI inflammation and activity index (J Immunotoxicology: Apr;5(2):179-87, 2008). This generic oral medication has now been tested in 2 adult clinical trials with Crohn's disease and one pilot trial in children. Due to its safety profile, low cost, and effectiveness, this drug should be considered as alternative treatment in those who do not respond to standard regimes.	Naltrexone was not a medication of interest.

Commentator & Affiliation	Section	Comment	Response
Public: Jill P. Smith, MD	Methods	Forty adult subjects with active Crohn's disease were evaluated in a double blind placebo controlled trial where they were randomized to naltrexone or placebo for 12 weeks. End points were response by CDAI scores and mucosal healing by colonoscopy and biopsies comparing pre to post-treatment.	Naltrexone was not a medication of interest.
Public: Jill P. Smith, MD	Results	More than 85% of the subjects in the naltrexone treated group had a clinical response and this was significant compared to controls (p=0.009). Also biopsies from those on naltrexone showed statistically significant mucosal healing while none on the placebo group were improved.	Naltrexone was not a medication of interest.
Public: Jill P. Smith, MD	Discussion	Alternaive therapies are needed for patients with inflammatory bowel disease that are safe, inexpensive and effective. Naltrexone is the first new class of drugs to be tested and shown to be efficacious in Crohn's disease since the biologics were approved. Efforts should be made to move the FDA to approve generic drugs like naltrexone for this new indication for IBD.	Naltrexone was not a medication of interest.
Public: Jill P. Smith, MD	Abbreviations	CDAI = Crohn's disease activity index IBD = inflammatory bowel disease	Thank you for pointing these out to us. We have confirmed that they are listed in the Abbreviations section of the report
Public: Jill P. Smith, MD	References	Am J Gastroenterology: 102:1-9, 2007 Dig Dis & Sci. 56:2088-2089, 2011.	Naltrexone was not a medication of interest.
TEP-2	General comments	The purpose of this exercise is to help health care providers, yet, the terminology used throughout is not clinically interpretable. "The strength of evidence was low to demonstrate no difference between x and y" has no clinical meaning. What is the difference between a low to moderate level of evidence to demonstrate no difference between x and y?	We provided a description of the approach to grading the strength of evidence in the methods section, with a definition of what each grade means. We also revised the discussion to provide a clinical perspective on the findings.
TEP-2	General comments	In several instances comparisons for "response" are mentioned. This is not a focus of this review and the inconsistent use of "response" should be avoided as it was not systematically assessed.	The references to response were removed from report.
TEP-2	General comments	At no point is definition of "moderate-severe" Crohn's disease mentioned.	References to moderate-to-severe disease have been removed from the results section of the report because studies generally didn't report results that way.

Commentator & Affiliation	Section	Comment	Response
TEP-2	General comments	Studies were included for induction of remission lasting 26-38 weeks. This is a misinterpretation of "induction" as the trials were designed to demonstrate "maintenance of a steroid-induced effect" (primarily thiopurines and methotrexate).	We agree. We are reframing KQ1 to emphasize that these patients have active disease at the time of randomization. For points after weeks 12-16 in KQ1, we are emphasizing this is maintenance of remission.
TEP-2	General comments	Maintenance trials should be at least 6 months to be included.	We agree. The first time point for KQ2 is 48-52 weeks
TEP-2	General comments	Key Question 4 assesses and irrelevant endpoint and only evaluated 1 trial which was NOT a post-operative prophylaxis study as it included patients enrolled 6-24 months after resection rather than within the 1st month.	According to our analytic framework, patient-reported outcomes are relevant. The trial met the eligibility criteria that we pre-specified.
TEP-2	Introduction; abstract	No mention of corticosteroids, aminosalicylates.	We mention corticosteroids and aminosalicylates in the introduction of the abstract.
TEP-2	Introduction; abstract	No mention of natalizumab as maintenance	Natalizumab is mentioned in the abstract for maintenance.
TEP-2	Executive Summary, Table A	Half life of aminosalicylates and sulfasalazine are not accurate or relevant pertaining to mechanism of action as their activity is luminal and not systemic. Mechanism of action is not known for aminosalicylates. Azathioprine is converted to mercaptopurine. Mechanisms of action are same (and unknown).	We have corrected this table.
TEP-2	Executive Summary	Why is purpose limited to moderate-severe Crohn's (vs. mild-moderate?). Budesonide is approved for mild-moderate by FDA. Moderate-severe is not defined.	We included all studies, not only those with moderate-to-severe Crohn's disease population.
TEP-2	Results, KQ1	Definition of remission does NOT include "a decrease in ….symptoms"	In the executive summary (ES-1), we have defined remission as patients who "no longer have active disease". We revised the report to be more consistent in our definition of remission.
TEP-2	Results, KQ1	Maintenance of remission at 12-16 &15-19 weeks has no meaning/interpretation. Least definition is 6 months. KO4 is based on 1 clinical study in 78 patients making entire topic not clinically interpretable.	We decided upon 1 year for our first maintenance endpoint
TEP-2	Executive Summary	Study characteristics included patients with CDAI 150-220…NOT CONSIDERED MODERATE TO SEVERE.	We agree. We decided we can't focus on moderate to severe disease and have included all activity levels.

Commentator & Affiliation	Section	Comment	Response
TEP-2	Executive Summary	"statistical significance was not required for an effect to be considered clinically significant"??? Then by what basis?	In the Data Analysis and Synthesis section of the Methods, we state, "For KQ1 and KQ2, we considered a difference to be clinically meaningful when there was a 10 percent absolute difference in the outcome between the groups compared, even when the difference was not statistically significant at a p-value less than 0.05."
TEP-2	Executive Summary	.."all activity levels and severities of Crohn's were included in most of the observational studies" hence, this does not apply specifically to moderate-severe disease. The risk/benefits then cannot be assessed for this category of disease severity.	We agree. We decided we can't focus on moderate to severe disease and have included all activity levels.
TEP-2	Executive Summary	Comparisons of intravenous azathioprine vs intravenous infliximab are irrelevant as former is not used in ANY clinical setting or practice.	We noted this in body of report and downgraded the strength of evidence for this reason.
TEP-2	Executive Summary	Again looking at only 1 trial in 78 patients for KQ4 makes this irrelevant.	We feel that patient-reported outcomes are relevant, regardless of the number of studies addressing it. The lack of studies addressing this question highlights it as a potential research gap.
TEP-2	Executive Summary	Strength of evidence for infliximab inducing improvement is not a criterion for this review.	We agree. We revised the report to focus on remission rather than response.
TEP-2	Executive Summary	Again, for certolizumab induces response is not relevant to topic.	We agree. We revised the report to focus on remission rather than response.
TEP-2	Executive Summary	strength of evidence was low that combination of ifx and aza was more effective than corticosteroids alone in inducing a steroid-free remission??? Makes no sense.	This means that the top down group was more effective than the step up group, but that the strength of evidence for this study was low. We have modified how the key points are written to be clearer.

Commentator & Affiliation	Section	Comment	Response
TEP-2	Executive Summary	"intravenous" methotrexate should read "intramuscular"	Thank you for pointing this out to us. We are referring to the Ardizzone 2003 study, and it reported using intravenous methotrexate.
TEP-2	Executive Summary	Many of the trials included a placebo arm that is not mentioned (Malchow, Summers, etc) and the table lists comparisons of prednisone vs. sulfasalazine.	Actually, we do include the placebo arms of these trials in the table.
TEP-2	Methods	My main critique is that "moderate to severe" Crohn's is not defined, nor followed in the analysis. Remission is not clearly defined or consistent in the analysis. The numerous time endpoints make it impossible to compare trials and too many time variables are included. Induction trials end at 16 weeks and maintenance trials at 6-12 month.	We removed "moderate to severe" from the Key Questions. We have provided definitions of remission throughout the report. We have specified in the Methods which time points we are including.
TEP-2	Results	The degree of detail makes any overall conclusions impossible (which is the conclusion of the study), that there are no conclusions. While one appreciates the effort at assimilating the tables and data...but the number of variables to the point of describing almost individual trials is not helpful. The tables are only useful as a reference. 99% of readers will read only the executive summary...which says nothing as conclusion is that there is no conclusive data.	We believe it is important to report the details in the body of the report. We have tried to revise the Executive Summary to present the findings and conclusions more clearly.
TEP-2	Discussion/ conclusion	The report was designed to be of use to clinicians and health care authorities. It is NOT useful.	We have included a section in the Discussion called a "Clinical Perspective."
TEP-2	Discussion/ conclusion	Moderate-severe Crohn's disease is not defined	We have removed references to moderate-to-severe Crohn's disease in the results section of the report.
TEP-2	Discussion/ conclusion	Remission is not defined	In the Introduction of the report, we provide this definition for remission, "Physicians refer to patients who no longer have inflammation as being in remission." We also created Figure 1, which illustrates remission vs. relapse.

Commentator & Affiliation	Section	Comment	Response
TEP-2	Discussion/ conclusion	Induction studies and maintenance studies are intermixed with overlapping endpoints	We have made it clearer that in KQ1 we are looking at patients with active disease at randomization, understanding that some of the longer trials are in fact looking at maintenance starting at around 12 weeks. We have edited the wording for KQ 1 to reflect that we are considering studies evaluated patients with active disease at randomization. We listed the timepoints that we are considering for KQ1 in Table B and in the Methods chapter under Data Analysis and Synthesis. Additionally, we provide our definition for KQ1 in the beginning of the KQ1 results section under Definitions.
TEP-2	Discussion/ conclusion	Terminology is not clinically useful (e.g. low strength of evidence for no difference?)	We have simplified the Key Points by separating the conclusion from the strength of evidence.
TEP-2	Discussion/ conclusion	Aside from no medicine or class is “most effective” and more research is needed, this report offers no guidance for anyone.	In the Discussion, we provided more details about the summary of key findings and added a clinical perspective.

Commentator & Affiliation	Section	Comment	Response
TEP-2	Discussion/ conclusion	Statement that statistical significance was not required for clinical effectiveness violates principles of evidence based medicine. In particular in absence of evidence for superior effectiveness.	Taken from Methods section (Data analysis and synthesis subheading): "For KQ1 and KQ2, we considered a difference to be clinically meaningful when there was a 10 percent absolute difference in the outcome between the groups compared, even when the difference was not statistically significant at a p-value less than 0.05. Similarly, we did not report statistically significant relationships unless there was a clinically meaningful difference."
TEP-2	Discussion/ conclusion	Effectiveness is not defined. Results are reported efficacy from clinical trials. Relative benefits versus relative risks are not compared or reviewed.	Throughout the report, we describe the results from clinical trials as "efficacy." In the Discussion chapter, we provide a section on "Clinical Perspective," which discusses the benefits and risks of different therapies.
TEP-2	Discussion/ conclusion	Executive summary and abstracts do not provide any recommendations or conclusions aside from "random" time points for individual agents (with poor separation of induction vs maintenance benefits).	In the Methods chapter under Data Analysis and Synthesis, we list the time points we considered in the review. In the Executive Summary, we provide tables that summarize the results at the specified time points. The evidence report is not intended to be a clinical practice guideline, but rather present the evidence to help clinicians and patients make better decisions about health care.

Commentator & Affiliation	Section	Comment	Response
TEP-2	Discussion/ conclusion	Comparison of intravenous azathioprine (which is not used at all!) versus infliximab is a completely irrelevant comparison. Discussions of intravenous azathioprine take a ridiculous amount of lines in summaries as this therapy is not used and offers no clinically useful comparison.	We removed mention of intravenous azathioprine from Executive Summary text because it is not used clinically. Results for intravenous azathioprine remain in main body of the report because we had no exclusions based on mode of administration.
TEP-2	Discussion/ conclusion	Aside from critique of the IBD community for needs of better comparative effectiveness trials this report offers no clinical guidance, whatsoever.	In the Discussion, we provided more details about the summary of key findings and added a clinical perspective.
TEP-2	Discussion/ conclusion	As stated in Executive summary, applicability for all key criteria are so limited as to have no utility.	We did not make any specific changes to the discussion in response to this comment.
TEP-2	Discussion/ conclusion	This reviewer considers this a failed attempt to compare effectiveness across agents in moderate-severe Crohn's disease.	We did not make any specific changes to the discussion in response to this comment.
TEP-2	Discussion/ conclusion	The future research is not helpful as it chastises the IBD community.	The Future Research Needs section was revised. It was not our intent to criticize the IBD community. Rather, we thought it was important to identify gaps in the evidence, and call for research that could fill the gaps.
TEP-2	Discussion/ conclusion	The authors "punted" on post-operative Crohn's by only assessing PROs and not clinical or endoscopic data	There was another recent, well-conducted systematic review on the effectiveness and safety of therapies for post-operative Crohn's disease, which is cited in the report. We decided to focus on patient-reported outcomes because these outcomes were not included in that systematic review. This decision was made during topic refinement and was agreed by the technical experts. We did not make any changes to the report in response to this comment.

Commentator & Affiliation	Section	Comment	Response
TEP-2	Discussion/ conclusion	There are no "main" points aside from the conclusion that the authors were unable to make conclusions using their methods. The incidence/prevalence is much smaller than other chronic diseases such as rheumatoid arthritis. The largest trials are pharma sponsored and not intended to demonstrate comparative effectiveness, rather to gain regulatory approval. Most smaller, investigator initiated trials had a limited scope.As mentioned, the methodology used and analysis precludes any utility for practice decisions or policy.	As indicated above, we revised the summary of key findings and added a section offering a clinical perspective.
TEP-2	Discussion/ conclusion	Back to the drawing board	We made extensive revisions in response to all constructive comments and suggestions.
Peer Reviewer-3	General comments	In this manuscript, Drs.Hutfless and Wilson have carried out a CER of pharmacologic therapies for the management of Crohn's disease. It is well written, comprehensive and methodologically of high quality. The goal of this systematic review is to examine the evidence that underlies treatment option where treatment variations between clinicians are potentially substantial. The authors carefully selected key questions by involving content experts, constructed an analytic framework detailing the effectiveness and safety questions in terms of final patient important outcomes. The search is comprehensive and the study selection process meticulous. They follow accepted methodological standards by using pre-defined tools (e.g., Cochrane risk of bias tool), meta-analytic techniques, and rating frameworks (e.g., GRADE framework as outlined for EPCs by Owens et al. JCE 2010).	Thank you!
Peer Reviewer-3	Results	Page 70, figure 2: please double check the study flow – the number of studies do not match up: 4868 minus 3857 does not equal 997; box "reason for exclusion at the abstract level…": sum is around 1000 less than the 3857 excluded abstracts; box "reasons for exclusions at the article level…" is a few hundred short of the 813 that were actually excluded.	This has been updated.
Peer Reviewer-3	Results	6 pages of key points: as this information is taken out of context, and the use of different approaches to convey effects (RD vs. RR vs. rates alone) and as this information is not repeated in the narrative section, there is too much information at once and it is very hard to try to find the information again when reading the individual sections. Would suggest to at least repeat the conclusions about the quality of evidence in the individual sections later in the text. The same applies to key question 2.	We now present the evidence grading tables in the text of the report as a summary of the findings.
Peer Reviewer-3	Results	I do not see the value in repeating too much study level detail in the text. For example: on page 92 (and following) studies are referred to as "underpowered". Instead, this should be discussed as an issue of precision and be based on the body of (pooled) evidence. Also, the text refers to clinical and statistical significance on a study level. Here again, the pooled summary estimates should be interpreted for imprecision in the light of clinical useful thresholds instead (e.g., see Guyatt et al.: GRADE guidelines 6. Rating the quality of evidence - imprecision. J ClinEpidemiol.	We provide study level detail when that information is important to interpret the results of the study (and the resulting body of evidence) and not provided elsewhere. We used GRADE formula to evaluate the strength of evidence, which included precision.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer-3	General comments	At 800+ pages, the entire report is very long (which, to a certain degree comes with comprehensiveness), but the usefulness to the end-user (guideline panels, policy makers, or even patients and clinicians) is quite limited by the convoluted way the data is presented. For example: Let's say I'm interested in finding out the comparative effectiveness of budesonide vs. regular steroids. I start reading the last paragraph on page 104 to find out that "there was no significant difference in the pooled remission rates…" RR 0.88 (CI 0.76, 1.02). However, there is no mention of quality of evidence. For this, I have to go back to page 74 where, after some searching, I can find the quality of evidence as "moderate" for no difference (however, the pooled RR is reported here as 0.9 (CI 0.8, 1.0) – different from page 104). Moderate quality means, that at some point the quality of evidence was rated down. So to find the reason, one need to go to evidence table 1 on page 395 in the appendix to find out that there was "medium risk of bias" across studies. However, there is no footnote stating what this judgment was based on (and table 5, page 531, does not necessarily help either, as 2 of the 3 studies that went into the analysis were deemed "good" and the remaining study (Campieri) was judged as fair due to unclear allocation concealment/sequence generation; no information is given how this influenced the overall judgment of quality of evidence across studies for this outcome).To be able to assess precision, I need to find the rates and number of events. Rates are found in the table 12 (page 107), but not the number of events, which I have to hunt down in the forest plot figure 5 (page 110) and use my calculator to sum it all up as the forest plot does not include the sum of events. Looking at the total number of events and the confidence interval that favors both budesonide and conventional steroids, precision seems clearly lacking (there is a worst case scenario of a 25% RRR for conventional steroid over budesonide) , but going back to table 1 (page 395) in the appendix there is again no documentation why precision was judged to be sufficient. In summary, the report would greatly improve with standard GRADE style evidence profiles, as finding, interpreting, and re-assessing the judgments made in this report is quite challenging across all outcomes.	We revised the order and tables as mentioned previously. Please see the end of the Search Results section for a description.
Peer Reviewer-3	Results	Referring to a table in the text that span 63 pages (e.g., table 4 in the appendix) is to a certain extent problematic, as it remains difficult to find the information.	The tables are already broken out by key question, and are labeled by type of drug. The summary tables in the main body of the report provide smaller snapshots of the data.
Peer Reviewer-3	Results	Under "Quality assessment": instead of stating: "…57% adequately generated their allocation sequence" I would suggest to change to "…57% reported adequately generated allocation sequence" as this is most of the time an issue of reporting (unless the trial authors were contacted (which was not reported) – then this statement does not need to be changed).	Agreed, edits made.
Peer Reviewer-3	Results	Page 120: under "Intervention": "The studies had significant heterogeneity". Please insert "clinical" before the word heterogeneity.	This edit was made.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer-3	Results	Page 124, line 36: RR at 48 weeks is listed as 0.7 (CI 0.6, 1.0). The same figures are used on page 150, but this refers to forest plot figure 11, which lists a summary estimate of 0.74 (CI 0.57, 0.97). If this is a rounding problem, I would strongly suggest not to round up or down, as for the statistically inclined an upper boundary of the CI of 0.97 is different than 1.0 (with the latter very close to being non-significant). Also, page 124 cites that the outcome at 48 weeks was based on 4 studies; however, figure 11 shows 3 studies. The evidence quality for the 5-ASA data is reported as low, due to imprecision and medium risk of bias. However, when looking at the evidence table 10 (page 631), it is somewhat difficult to understand how this was concluded: Mahmud 2001 was judged as good, as was Prantera 1992 despite the unclear allocation sequence and concealment; Sutherland 1997 was judged as poor, although there did not appear any methodological flaws listed. In general, it is unclear whether in the column heading "incomplete outcome data" a "yes" means incomplete data reporting (= poorer quality) or "yes" means complete data (looking at the overall quality statements, this is probably what is meant, so this is confusing).	We have removed the pooled relative risk from the Key Points and removed the meta-analysis.
Peer Reviewer-3	Results	Not sure it is informative to list a single case of PML associated with natalizumab. PML does not occur unless patient is profoundly immunosuppressed (historically seen in HIV and cancer patients). I would strongly suggest reporting all cases of PML associated with natalizumab regardless of indication (MS, IBD) and set the number in relationship to the estimated total number of treated patients to better illustrate the risk (e.g., 31 cases of PML associated with natalizumab: 0.3 to 0.8 per 1,000 patients; from http://www.fda.gov/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/ucm199872.htm).	This is outside the scope of our project. No changes made.
Peer Reviewer-3	Appendix C	Other minor issues: Appendix C: list of excluded articles appears to be wrong, as this list includes multiple studies that were actually included in the analysis (e.g., Bar-Meir- page 334); it seems the reason for exclusion "no subject with Crohn's disease" is frequently wrong.	These studies were excluded because the patient population was not exclusively Crohn's disease. We have checked that this is a correct exclusion for several of these studies. The reason for exclusion for these studies has been edited to reflect this. The Bar-Meir study should not have been listed as an excluded article, and has been removed from the excluded article report.