Comparative Effectiveness Research Review Disposition of Comments Report

Research Review Title: *Allergen-Specific Immunotherapy for the Treatment of Allergic Rhinoconjunctivitis and/or Asthma*

Draft review available for public comment from April 27, 2012 to June 14, 2012.

Research Review Citation: Lin SY, Erekosima N, Suarez-Cuervo C, Ramanathan M, Kim JM, Ward D, Chelladurai Y, Segal JB. Allergen-Specific Immunotherapy for the Treatment of Allergic Rhinoconjunctivitis and/or Asthma: Comparative Effectiveness Review. Comparative Effectiveness Review No. 111. (Prepared by the Johns Hopkins University Evidence-based Practice Center under Contract No.290-2007-10061-I.) AHRQ Publication No. 13-EHC061-EF. Rockville, MD: Agency for Healthcare Research and Quality. March 2013. Errata added May 2013. www.effectivehealthcare.ahrq.gov/reports/final.cfm.

Comments to Research Review

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

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

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

Source: http://effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=1427
Published Online: March 27, 2013

Comment #	Commentator & Affiliation	Section	Comment	Response
1	Peer Reviewer #1	Executive Summary	There is quite a bit of detail. The executive summary is excellent, so readers have a choice whether to explore the detail or read the salient features of the review.	Thank you for your comment
2	Peer Reviewer #1	Introduction	well written	Thank you for your comment
3	Peer Reviewer #3	Introduction	Good summary.	Thank you for your comment
4	Peer Reviewer #2	Introduction	The Introduction section includes an appropriate description of the prevalence of allergic rhinoconjunctivits and asthma. The 2 forms of immunotherapy that are being considered in this report – subcutaneous and sublingual immunotherapy – are introduced and described. The rationale for this comparative effectiveness review is noted.	Thank you for your comment
5	Peer Reviewer #1	Methods	The methods are well written, easy to understand and appropriate.	Thank you for your comment
6	Peer Reviewer #2	Methods	The inclusion and exclusion criteria are appropriate and justifiable. The grammar and context in which the inclusion/exclusion criteria are stated in Table 1 could be refined – consider removing first person "We…" in the inclusion/exclusion criteria statements.	Thank you for your comment; we consider it is OK to leave it as it is.
7	Peer Reviewer #2	Methods	The search strategies are logical and well stated.	Thank you for your comment
8	Peer Reviewer #2	Methods	The description of bias assessment is clear and appropriate.	Thank you for your comment
9	Peer Reviewer #2	Methods	The numbers for included/excluded studies are inconsistent across multiple sites in the text/figures. For example, the number and breakdown of included articles in Figure 3 is not consistent with what is stated in the Results paragraph on page 12 (lines 45-49) and page 44 (lines 5-9). Please ensure that all numbers are consistent and appropriately additive across all sections/figures, or provide an explanation for inconsistencies.	Thank you for your comment, we revised all the sections in the report to make sure the numbers were consistent and corrected the mistakes

Comment #	Commentator & Affiliation	Section	Comment	Response
10	Peer Reviewer #2	Results	The number of included SLIT studies stated in the text on page 44 (lines 47-49) does not coincide with the graphical representation in Figure 5. The text notes that the majority of SLIT studies enrolled at least 100 subjects, but the chart indicates that the majority of SLIT studies enrolled at least 50 subjects.	Thank you for your comment, We noted the mistakes and corrected them
11	Peer Reviewer #2	Results	The detail in the Results section is clear and appropriate. The characteristics of the studies are clearly delineated.	Thank you for your comment
12	Peer Reviewer #2	Results	The report is organized according to Key Questions amongst categories of immunotherapy modality and adult vs. pediatric populations. This is intuitive and relatively easy to follow.	Thank you for your comment
13	Peer Reviewer #2	Results	It appears that the authors have included appropriate studies in this report. The authors have done a nice job of grading evidence where it is appropriate and providing descriptive summaries when evidence grading was not applicable.	Thank you for your comment
14	Peer Reviewer #4	Results	Report Weaknesses: Perhaps the most limiting weakness of this report is that its body of evidence is based on published literature, alone. This means that no weight is given to unpublished studies. Further, small studies [Phase I (safety); Phase II (dose-ranging; mechanistic)] may show effects which cannot, and are not reproduced in larger field trials (Phase IIB/Phase III), and hence never get published.	As discussed in the methods and in the results sections, we requested the Information Packages from the pharmaceutical and manufacturing companies to reduce the of risk bias from not including gray literature in the analysis. We did not receive anything from them.

Comment #	Commentator & Affiliation	Section	Comment	Response
15	Peer Reviewer #4	Results	a) *Incorrect Interpretation of Findings:* A perfect example of this dilemma is that of the work performed in the U.S. on oral aqueous ragweed. The report cites the study by Skoner (reference #: 131) of 115 ragweed-allergic patients with RCS. The findings from this study were not significant for the primary endpoint [only a 15% reduction in symptom scores compared to placebo during the entire pollen season (p: > 0.1)] in the Intent-To-Treat (ITT) population. [Furthermore, in the Discussion section, it is borne out that medication use, likewise, did not reach significance for the SLIT-treated group]. Only when subgroup analyses were performed on the data, was it possible to show an effect on symptoms or medication scores (See Table II)].	The methodology to assess the body of evidence is described in the methods section. Two reviewers independently assessed the risk of bias in each article and came to consensus about the overall rating. We used a modification of the Cochrane Collaboration Tool for Assessing Risk of Bias from the Cochrane Handbook for Systematic Reviews of Interventions. This tool was used to assess potential sources of bias: 1. Was there random allocation of subjects? 2. Was the allocation scheme concealed? 3. Was the intervention concealed from study personnel and participants? 4. Was incomplete data adequately addressed? 5. Were there other important sources of bias? While the reviewer raises concerns about study limitations, we followed the methods described, and assessed this particular article to fall into an overall low risk of bias. (Please see table E4- Appendix E to find the quality assessment of this study) We disagree with the reviewer as this does not represent a sub-group analysis. This paper did had 3 arms in this study and per Skoner's article, the patients were randomized into 3 groups (higher dose, lower dose, and placebo) at the very beginning of the study. The Skoner article's objective is to determine a safe and effective maintenance dose, by comparing placebo to 4.8 microgram group (low dose) and 48 microgram group).

Comment #	Commentator & Affiliation	Section	Comment	Response
15	Peer Reviewer #4	Results		In our SR the results of grading do not necessarily correspond with the statistical findings for each study. **Our group did not use statistical significance/precision to grade articles.** We could not comment on the precision of the effect sizes as there were seldom measures of variance within the individual studies. (This is explained in pages ES-6 and in the methods section, page 10) We did not use the reported statistical significance of the differences between groups to grade the evidence as this was rarely reported. Instead, as described in our methods section, In our grade assignments, we considered the limitations of each individual study's quality (using the risk of bias classification), the consistency of the direction of the effect across studies, the directness of the body of evidence to the question of interest, and the magnitude of the effects reported across trials. While the reviewer is correct to say the Skoner article describes a 15% reduction in rhinoconjunctivitis scores over the whole season (seasonal average), that is not the method we utilized in our systematic review. In our review, we used raw data included in the papers to determine the percent change in outcomes in the intervention arm, and also the percent change in the comparator arm; the magnitude of effect was based on the difference between comparators. For studies such as this Skoner article that reports one season of treatment, the raw data values that would be used per our methods would include pre treatment scores versus peak season scores, and this data was not available in the paper for rhinoconjunctivitis scores so we could not determine magnitude of effect for this particular outcome.
15	Peer Reviewer #4	Results		The Skoner article evaluated two different outcomes, Rhinitis/rhinoconjunctivitis symptoms scores (See report Page 57 and table 27 in page 59) for which the magnitude of effect could not be determined and Medication scores (See report Page 61 and table 29 in page 63) in which we found a strong magnitude of effect for the High dose and moderate for the low dose. Specifically for the primary outcome of rhinoconjuntivitis symptoms scores, the Skoner article (as demonstrated in the Body of Evidence table) was graded as low risk of bias, positive direction of change, direct, but the magnitude of effect could not be determined from the data presented in the paper. However, for medication scores, the higher dose group had a strong magnitude (>40%) of effect, and the lower dose group a moderate (15-40%) magnitude of effect.

Comment #	Commentator & Affiliation	Section	Comment	Response
16	Peer Reviewer #4	Results	In this context, the Intent-To-Treat (ITT) analysis is the accepted primary method of statistical analysis for evaluation of a clinical trial. An acceptable alternative in some regulatory circles is the Per-Protocol (PP) analysis. However, subgroup analyses are best-recognized as "exploratory" parameters that provide insight into future study designs (and which need to be "confirmed" in such subsequent clinical trials). [See WAO Taskforce Report, cited below]. Indeed, these type analyses are often performed post-fact in an attempt to "mine the data" to learn as much as possible about a study drug's effects, to better understand "failings" in a given study, and to optimize subsequent study design.	We agree, Intent to treat analysis is an accepted method for the evaluation of individual clinical studies. We used the GRADE tool, as a recognized method when performing effectiveness reviews. See the Grading of Recommendation Assessment, Development and Evaluation (GRADE) Working Group, adapted by AHRQ in the Methods Guide for Effectiveness and Comparative Effectiveness Reviews (http://www.effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productid=328) For this particular article, Skoner had 2 groups in the ITT; the purpose of their article was to find a safe and effective dose and therefore designed to look at a lower dose group, and a higher dose group; these 2 treatment groups were compared to a 3rd placebo group. Skoner clearly describes these groups in his methods. Our review included data on both these ITT groups as was presented in the original article by Skoner. Our review did not do any further subgroup analysis. In addition, on reviewing Skoner's original paper, we did not find any evidence that Skoner performed any post-fact analyses or manipulation of the data that had not been clearly outlined and consistent with the methods stated in his study.

Comment #	Commentator & Affiliation	Section	Comment	Response
17	Peer Reviewer #4	Results	I would note, the findings from this study would be difficult to interpret as "strong" *magnitude of effect* [page 88; line 29-30]. In fact, it should be considered "weak" based on the statistical methods (subgroup analyses) which were required to demonstrate "positive" effects in a subset of study subjects. Furthermore, a 15% reduction in symptoms in the "active" treatment group would not be considered clinically meaningful with immunotherapy (even if it were statistically significant; which it is not).	The methodology to assess the body of evidence is described in the methods section. In our SR the results of grading do not necessarily correspond with the statistical findings for each study. As described in the methods section, every study was individually assessed and the evidence was rated according to the individual quality of the studies (Risk of Bias tables), consistency, directness and magnitude. Magnitude of effect should not be confused with Strength of evidence. As described in our methods section: Strength of Evidence is determined by grading the overall body of evidence. In our grade assignments, we considered the limitations of each individual study's quality (using the risk of bias classification), the consistency of the direction of the effect across studies, the directness of the body of evidence to the question of interest, and the magnitude of the effects reported across trials. Magnitude of effect was determined as follows: We calculated the percent change in outcomes in the intervention arm, and also the percent change in the comparator arm; the magnitude of effect was based on the difference between comparators. Magnitude of effect was classified as weak if there was less than a 15 percent difference in percent change between the SIT group and comparator arm; a 15 to 40 percent difference was called moderate, and greater than 40 percent was considered a strong effect. For studies that recorded outcomes at multiple time points, we used the outcome data from the final time point reported and compared to pre-treatment values. However, some studies treated and assessed subjects for only one season; in these single season studies, the values reported at peak pollen seasons were used when available.

Comment #	Commentator & Affiliation	Section	Comment	Response
18	Peer Reviewer #4	Results	The European recommendations for standardization of clinical trials [Recommendations for Standardization of Clinical Trials with Allergen Specific Immunotherapy for Respiratory Allergy: A Statement of a World Allergy Organization (WAO) Taskforce; GW Canonica, CE Baena-Cagnani, PJ Bousquet, et al; Allergy 2007; 62: 317-324] cites that "a minimally clinically relevant efficacy should be at least 20% higher than placebo". The same concensus is agreed upon by research investigators in the U.S./N. America.	While Canonica et al refer to placebo-controlled studies, our review included studies comparing SIT to pharmacotherapy, and for these studies we would expect less difference. Hence, we believe that a 15% difference is acceptable. (This is explained in page ES-14 and in page 10 of the report) That is one number in one review, other reviews find different numbers to determine relevant efficacy. We calculated the numbers from the raw data available in each study. The magnitude of effect in a trial was classified as "weak" if there was less than a 15 percent difference in post-to-pre change comparing the SIT group and comparator group, 15 to 40 percent difference was called "moderate", and greater than 40 percent was considered a "strong" effect. These were set after a consensus decision as to what were felt to be clinically significant percentages that would affect clinical decision making. Other TEP members and Peer reviewers agreed with this number.
19	Peer Reviewer #4	Results	The interpretation of any study, but in particular, this study, requires a reviewer that has an understanding of study design methodology, a knowledge base to interpret clinical outcomes, and most importantly the ability to correctly interpret the statistical methods used in the analyses performed.	Thank you for your comment
20	Peer Reviewer #4	Results	b) *Lack of Non-published Data* As a result of this partial evidence of efficacy from this small clinical trial, the pharmaceutical company pursuing this research received guidance from the regulatory authorities to perform a larger study.	As discussed in the methods and in the results sections, we requested the Information Packages from the pharmaceutical and manufacturing companies to reduce the risk bias from not including gray literature in the document. We did not receive anything from them.

Comment #	Commentator & Affiliation	Section	Comment	Response
21	Peer Reviewer #4	Results	This next clinical trial was designed based on what had been learned from this initial Phase II (dose-ranging) clinical trial and it was anticipated that with a larger study subject enrollment, selection of a proper dose, and refinement of study methods to reduce patient variability, that a successful clinical trial could be performed. However, this subsequent Phase IIB/III clinical study, of 400+ ragweed-allergic patients with RCS, completely failed to demonstrate efficacy (2010 press release: Greer announced that ragweed clinical trial failed to demonstrate positive clinical findings). This negative study has not been published. This clearly shows the faulty conclusions that can be generated based simply relying only on published data to provide "evidence for effect". It also clearly shows the (unanticipated) inherent "bias" of the report's findings, as a result of not appreciating the total body of clinical work that is in the public domain.	We acknowledge the possibility of missing unpublished data which raises concern for publication bias. As discussed in the methods and in the results sections, we requested the Information Packages from the pharmaceutical and manufacturing companies to reduce the of risk bias from not including gray literature in the analysis. We did not receive anything from them so we did not know about this study. We also searched Clinicaltrials.gov seeking for the literature resulted from finalized or ongoing clinical trials. All the references we found were already included in our database. This limitation is clearly stated in our discussion section- page 108
22	Peer Reviewer #4	Results	The corollary to this discussion is that what remains true in the U.S. is that no strong body of work exists to provide "strength of evidence" for approval of SLIT. The predominantly European body of literature is not to be overlooked; in fact, its weight of evidence is moderate-to-strong for their reported published studies. However, one of the main goals of this CER report is to provide the U.S. physician with quality information on his/her consideration for using SLIT in as an off-label drug. The interpretation of finding from the selected clinical trials must be prefaced with the understanding that data from U.S. studies is not only weak but un-supporting.	We do not agree with this comment. The goal of the CER is to provide the reader with quality, up to date and impartial information regarding the use of immunotherapy. We analyzed the existent data without any prejudice of where it came from. It is not the goal of the CER to brand the quality of the U.S studies, but to interpret the data. Our review included European SLIT studies and also U.S. studies (4 studies). Table E1-Appendix E. First, the number of U.S. studies is less than 7% of the total number of studies so is a very small part of the total articles. Furthermore, our CER did not find significant differences between the European data and the U.S data, therefore we do not believe U.S data should be presented separately.

Comment #	Commentator & Affiliation	Section	Comment	Response
23	Peer Reviewer #4	Results	Within this framework, there is some data from U.S. studies in which subgroup analyses performed and showed, in small numbers of subjects, positive changes in subsets of patients (stratified based on greater clinical sensitivity or selected biomarkers) – but this simply points out the shortcomings of "said" studies, and the need to perform properly powered trials with appropriately allergic patients and correct methodology.	We agree with this reviewer and we address this limitation and need in our Future Research Needs section (page 111-paragraph 3). Studies are not adequately powered to assess specific subgroups.
24	Peer Reviewer #4	Results	The real problem now is to determine why this is the case -- that is, what are the factors that have exposed the differing results between the U.S. and European data. In this context, geographic factors (intensity of pollen exposure; confounding allergens; topography), degree of patient sensitivity (single vs multiple allergen; genetic factors; level of IgE; cytokine milieu), environmental interactions (pollution; urban vs rural; smoking; etc), and other confounding disease (non-allergic rhinitis; sinusitis; asthma) are just a few of the important factors that can impact upon a therapeutic effect with immunotherapy.	Our review included all studies that met our inclusion criteria: we included studies from Europe, Asia, North America (U.S. and Canada), South America and Australia. Our goal was to analyze the data without any prejudice of where it came from. From the 142 studies included, only 16 studies (11%) came from the U.S.; 12 studies in the SCIT group, 4 in the SLIT group and none in the SLIT vs SCIT group. The number of U.S. studies is such a small part of the total articles and the authors did not find significant difference, therefore we do not believe U.S data should be presented separately. There were very few studies addressing the mentioned subgroups and we do not think the heterogeneity is derived from the origin of the studies but by the literature itself.
25	Peer Reviewer #4	Results	<u>Allergen Strength of Evidence</u> The Executive Summary (ES) section (page ES-7; ES-8) provides a concise overview of the finding on KQ#1; however, the summarization is not necessarily consistent with the Body of Evidence: a) For instance, on page 24 [Table 2: Body of Evidence for Subcutaneous Immunotherapy and Asthma Symptom Scores], the Table cites 15 clinical studies (13 with single allergen treatment; 2 with multiple allergens). A "magnitude of effect" score is cited for each study (varying from *weak* to *strong*, depending on the allergen used for treatment).	We revised all numbers in the text and the ES and corrected the mistakes

Comment #	Commentator & Affiliation	Section	Comment	Response
26	Peer Reviewer #4	Results	However, in the ES (ES-7), it is stated that the strength of evidence is high that subcutaneous immunotherapy improves asthma symptoms – clearly, this "effect score" is influenced by the treatment allergen (seasonal vs perennial) as well as the allergen mixture (single vs multiple).	Since most of the studies use single allergens, the statements in the text and the ES tables reflect the outcomes for single allergens. We added a statement in the text to avoid confusion.
27	Peer Reviewer #4	Results	We would be better served, if it was possible, to assess the "magnitude of effect" based on the respective categorization of allergens (dust mites; pollens; animals; mold) – of course, it is clear that an insufficient number of studies have been identified from the literature (that meet the criteria) except for dust mites and pollen. But that, in and of itself, is an important observation (and statement to make) and makes it difficult to provide a "blanket statement" with respect to the ES statement.	Sub categorization of the results by allergens is not a relevant task, since the data very limited to draw conclusions on subgroups. However, we added a sentence on each outcome, pointing to which was the most relevant or frequent allergen for each outcome.
28	Peer Reviewer #4	Results	Furthermore, as there are only two cited multiple allergen studies, and the results are conflicting, it would make me less inclined to incorporate these study results into the single allergen study findings.	We did not incorporate the multiple allergen results in the single allergen results, in the text you can find the articles discussed as separate entities.
29	Peer Reviewer #4	Results	In addition, having reviewed the Adkinson study findings, and personally having talked directly with Dr. Adkinson, I find it hard for this paper to have been graded as meeting the criteria for *moderate* "magnitude of effect". [In fact, of the multiple endpoints assessed (>20), none were statistically significant (in favor of SCIT), except for the secondary PEFR finding (p: 0.05)]. This either reflects qualitative misjudgment or incorrect interpretation of the paper's findings. This is one example of this discrepancy in interpretation of findings, so it will be necessary to address this collectively with respect to the various Tables in the report.	For asthma symptom scores, this paper was graded as having moderate magnitude of NEGATIVE effect. The direction of change was negative – indeed supporting the reviewer's comments that this finding was NOT in favor of SCIT. The magnitude of this negative effect was moderate because it exceeded the threshold for weak. Statistical significance had no role in determining the direction and magnitude of effect. For asthma medication scores, this paper was graded as having weak magnitude of positive effect because it met the threshold for weak, without achieving statistical significance.

Comment #	Commentator & Affiliation	Section	Comment	Response
30	Peer Reviewer #4	Results	b) The same issues and concerns are noted with the SLIT conclusions. For example, the "body of evidence" Table [page 56; Table # 27; incorrectly labeled Table #33) shows 35 studies [weak: 16; moderate: 5; strong: 13; can't determine: 1]. [Query: The text refers to 36 studies, but the Table line-items 35 studies (with O'Hehir repeated twice in its box) and Valoverta having a Savolaine included in his box (but no data).	We revised all numbers in the text and the ES and corrected the mistakes
31	Peer Reviewer #4	Results	[page 56; Table # 27; incorrectly labeled Table #33)	All the labeling and numbering of tables has been reviewed and corrections have been made when needed
32	Peer Reviewer #4	Results	For this SLIT Tablet (and the above SCIT Table), it would be much better to distinguish the findings based on the specific allergen employed for treatment: (See table in the Document) This type handling of the data would certainly allow the reader to have a clearer understanding of the evidence supporting SCIT or SLIT for a given allergen.	All the grading tables are organized by allergen type. However some outcomes tables have repeated allergens and some do not. We appreciate your example table but we don't think it is helpful. We do not think the effectiveness of SLIT and SCIT is dependent on the specific allergen. Rather, our findings reflect the fact that some allergens have been more extensively studied, and as a result more data exists for these allergens than for other allergens.
33	Peer Reviewer #4	Results	c) One of my concerns is that the specific studies being graded aren't being properly assessed. For instance, the Nelson (cat) study was not able to demonstrate a clinical benefit vs placebo (NS); in this study a strong placebo effect was noted (and this possibly contributed to the study's "lack of effect"). Nonetheless, the primary clinical endpoint was not significant – therefore, it is hard to understand the grade of "strong" magnitude of effect.	In our SR the results of grading do not necessarily correspond with the statistical findings for each study. As described in the methods section, every study was individually assessed and the evidence was rated according to the individual quality of the studies (Risk of Bias tables), consistency, directness and magnitude. Clarification: In our SR, there is an assessment of the effect (direction, magnitude) and an assessment of the confidence in that effect. The confidence is based on the strength of evidence as determined by the risk of bias of the studies, consistency, directness and magnitude. This particular study was assessed for rhinoconjunctivitis symptoms as follows: medium risk of bias, positive direction of change, direct, weak magnitude of effect. We reviewed the data and found that for this specific study the magnitude of effect needed to be changed. However, this did not change the SOE.
34	Peer Reviewer #4	Results	d) A number of these studies are "open-label"; therefore, their "demonstrated" clinical findings are difficult to interpret.	As discussed in the results section, 72 percent of the studies are double-blind RCTs, we think the findings are interpretable.

Comment #	Commentator & Affiliation	Section	Comment	Response
35	Peer Reviewer #4	Results	Other Comments: a) I think that it is unfortunate that objective measures of immunotherapy's clinical effect (skin titration/nasal provocation/bronchial provocation/PFTs/PEFR) were not graded. These secondary measures of clinical efficacy are important outcomes that further define the therapeutic effect with this drug class. Likewise, changes in IgG vs IgE are predictors of successful outcomes with immunotherapy.	We did not grade Pulmonary function test because it is an intermediate outcome. We explained the reason in the methods section. See Owens et al. AHRQ Series Paper 5: Grading the strength of a body of evidence when comparing medical interventions Agency for Healthcare Research and Quality and the Effective Health-Care Program. J Clin Epidemiol, 2010;63(5):512-23.
36	Peer Reviewer #4	Results	b) Sub-categorization of effect based on dose would be a useful determinant to ascertain therapeutic effect, as this is a critical determinant in successful outcomes with immunotherapy (both for SCIT and SLIT). It would be interesting to see the "strength of evidence", when the trials are separated by high vs low dose or defined levels of dose.	We agree, but the articles comparing doses were very few to analyze. We have a paragraph in the discussion (Future Research Needs) about the need for studies comparing different regimes (duration, doses and dosing strategies)
37	Peer Reviewer #4	Results	c) The distinction with respect to multiple allergens is important. Very few studies include "multiple" allergens (grass + weeds + trees; grass + house dust mites + cat; etc). The term "multiple" allergens, can be confusing (or actually deceptive). For instance, the Temperate grasses cross-react heavily (Timothy/Orchard/June/Fescue); therefore, it is wrong to use this "multiple" allergen mix to support evidence for multiple allergens being effective in immunotherapy.	All the allergens are described very carefully in the appendices. Several authors described the allergen as "mix . We used the term mix as it was used in the articles
38	Peer Reviewer #1	Discussion/ Conclusion	All implications are clearly stated; no important literature is omitted. The conclusions identify issues that should be studied that would add to the knowledge base.	Thank you for your comment
39	Peer Reviewer #1	Discussion/ Conclusion	A few suggestions for the manuscript (different sections): 1. The reviewers may wish to comment that the U.S. SLIT data may not accurately address the efficacy of SLIT because tablets or preparations intended for that use would allow a higher quantity or concentration of allergen to be used than current licensed preparations allow.	The doses used in the U.S. studies were similar in micrograms of major allergen to other non-U.S. studies (please see Appendix E, Table 14, which gives doses in micrograms of major allergen per month for comparison where possible. Clarification: Allergen preparations licensed for SCIT are being used in the United Stated in an off-label manner for SLIT and with similar doses (in micrograms) seen in studies that have been performed in Europe and elsewhere (See Table E4-Appendix E). Also presented in our discussion section.

Comment #	Commentator & Affiliation	Section	Comment	Response
40	Peer Reviewer #1	Discussion/ Conclusion	2. The authors classify changes (pre- vs. post immunotherapy) less than 15% differences as "weak." That is appropriate. The authors may wish to state, however, that changes of 15% may reflect a very real improvement in quality of life for some patients.	This is added to the discussion under applicability in the executive summary.
41	Peer Reviewer #1	Discussion/ Conclusion	3. Studies addressing the efficacy of SIT in asthmatics may demonstrate efficacy for mild asthmatics but not severe asthmatics because dosage may be limited in the latter due to life threatening side effects. This may also be the case for studies of multiple allergens.	While we agree that this is a reasonable comment, we can only say based on our review that "findings from a few studies support that subcutaneous immunotherapy is more beneficial in patients with mild asthma than with severe asthma." We have discussed this in the Applicability section. (pages 105-106)
42	Peer Reviewer #1	Discussion/ Conclusion	Finally, and perhaps most importantly, this reviewer feels strongly that it should be stated in the executive summary that although SLIT appears to be much safer than subcutaneous immunotherapy, there are few studies of SLIT in moderate or severe asthmatics, and the safety data in mild asthmatics or those with allergic rhinitis must not be extrapolated to the more severely affected patients. This reviewer is concerned about off label use of licensed products for SLIT by physicians who are falsely comforted by the current safety data.	Thank you for your comment, this was already in our discussion section. To make it clearer we added a statement about this to the executive summary under limitations (Page ES-15), we also discussed in the full report. We have moved this discussion to the applicability section (page 105-last paragraph).
43	Peer Reviewer #3	Discussion/ Conclusion	The report clearly identifies unanswered issues.	Thank you for your comment
44	Peer Reviewer #2	Discussion/ Conclusion	The major findings of this review are clearly stated in the Discussion section. The limitations are well described and closely mirror the limitations of prior systematic reviews and meta-analyses on this subject. The Future Directions section appropriately delineates research needs.	Thank you for your comment
45	Peer Reviewer #2	Figures	The figures and tables are used appropriately. They are illustrative and descriptive of the major points of the report. The lists of studies included in the tables within the Results section do not seem to follow any particular order. It would be helpful to organize the studies in some fashion – chronologically within antigen categories, for example.	Thank you for your comment

Comment #	Commentator & Affiliation	Section	Comment	Response
46	Peer Reviewer #2	Figures	The lists of studies included in the tables within the Results section do not seem to follow any particular order. It would be helpful to organize the studies in some fashion – chronologically within antigen categories, for example	The studies are organized by allergen, and within allergen the placebo controlled studies go first.
47	Peer Reviewer #1	General	Very well written report; comprehensive and valuable review of the data.	Thank you for your comment
48	Peer Reviewer #3	General	The report is an excellent compilation and analysis of available data. Key questions and audiences are identified and are appropriate.	Thank you for your comment
49	Peer Reviewer #2	General	This report presents a systematic review of randomized controlled trials of subcutaneous and sublingual immunotherapy. The review was designed to evaluate the clinical efficacy, effectiveness, and safety of these immunotherapy modalities versus placebo, pharmacotherapy, or other forms of specific immunotherapy. As the review contained only randomized controlled trials, real world effectiveness does not seem to be specifically evaluated in this report, per se. The authors do not really tease this issue out (as being separate from efficacy within the context of a controlled trial)	Thank you for your comment. We have discussed this point under limitations.
50	Peer Reviewer #2	General	Given the high prevalence of allergic rhinoconjunctivitis and asthma in the US, this report is of substantial clinical importance and the data it contains are meaningful in the context of appropriately selecting treatment methods in an allergy practice.	Thank you for your comment
51	Peer Reviewer #2	General	The target audiences for this document are defined as clinicians providing care for allergic patients, allergic patients making decisions regarding specific immunotherapy treatment, and guideline developers who will make recommendations about the use of allergen immunotherapy.	Thank you for your comment
52	Peer Reviewer #2	General	The 3 Key Questions are nicely stated, straightforward, and testable.	Thank you for your comment

Comment #	Commentator & Affiliation	Section	Comment	Response
53	Peer Reviewer #2	General	There are numerous grammatical/typographical/formatting errors and inconsistencies in the text, tables, and appendices. Close editing would be beneficial in reducing these distractions.	The document was revised and edited before resubmission
54	Peer Reviewer #4	General	The conclusions for SCIT and SLIT need to be clearly stated in terms of their separate findings. The mandate of this AHRQ report is to provide evidence for the effect and safety of allergen immunotherapy products in the U.S. It needs to be cited "upfront" that SCIT is FDA-approved for use by physicians in the U.S.; whereas, SLIT is currently being used off-label in the U.S.	The document is clearly divided SLIT and SCIT as separated chapters. We have a statement in the background of the ES and the introduction asserting the off-label use of SLIT in the US.
55	Peer Reviewer #4	General	Sample Language for Executive Report: The report should state: "the evidence for SLIT is based upon a preponderance of properly controlled European clinical trials that demonstrated clinical efficacy and a sufficient degree of safety to justify regulatory approval in Europe. However, studies performed in the U.S., to date, have not met these criteria, and further studies are warranted, and being conducted. As a result, the use of SLIT in the U.S. is 'off-label'".	Thank you for your suggestion. We have a statement in the background of the ES and the introduction asserting the off-label use of SLIT in the US.
56	Peer Reviewer #4	General	The implicated conclusion should not be that these two forms of treatment are synonymous. In this context, there are several places in the report where it can be implied (from overall strength of evidence) that the findings for SCIT and SLIT are synonymous, yet this is not borne-out by the data.	We do not agree with this comment. We do not think we are implying the synonymy of both treatments. SLIT and SCIT are analyzed and discussed in separate chapters and the conclusion discusses the findings as the two different entities they are.
57	Peer Reviewer #1	Clarity and usability	Very clear	Thank you for your comment
58	Peer Reviewer #3	Clarity and usability	The report is highly useful and conclusions are clear.	Thank you for your comment
59	Peer Reviewer #2	Clarity and usability	In general, the report follows a clear and organized structure. The main points are highlighted, and the descriptions are relatively concise.	Thank you for your comment