Minimal Clinically Important Difference (MCID) in Allergic Rhinitis: Agency for Healthcare Research and Quality or Anchor-Based Thresholds?

Eli O. Meltzer, MD, Dana Wallace, MD, Mark Dykewicz, MD, and Lucy Shneyer, MA
San Diego, Calif; Fort Lauderdale, Fla; St. Louis, Mo; and Denville, NJ

BACKGROUND: In 2013, the Agency for Healthcare Research and Quality (AHRQ) in the USA recommended using an MCID equal to 30% of the maximum TNSS as a useful threshold. Treatment differences that failed this threshold would indicate equivalence. However, evaluations testing this threshold by the AHRQ and subsequent investigators could not demonstrate differences in effectiveness between various treatments for seasonal allergic rhinitis.

What is already known about this topic? Multiple approaches have been suggested for estimating a minimal clinically important difference (MCID) for allergic rhinitis studies, with most based on the total nasal symptom score (TNSS). Most recently, in 2013, the Agency for Healthcare Research and Quality (AHRQ) in the USA recommended using an MCID equal to 30% of the maximum TNSS as a useful threshold. Treatment differences that failed this threshold would indicate equivalence. However, evaluations testing this threshold by the AHRQ and subsequent investigators could not demonstrate differences in effectiveness between various treatments for seasonal allergic rhinitis.

What does this article add to our knowledge? This article describes the use of a threshold determined using a validated anchor-based approach that can be applied to allergic rhinitis clinical studies with appropriate data. By applying this threshold to 3 of the queries in the AHRQ report, using the same database, the article demonstrates the differences in outcomes. MCIDs for patient symptom relief were attainable for the majority of studies, despite the negative results reported by the AHRQ. In contrast to the results of the AHRQ analysis, the outcomes shown in this article are those that would be expected based on other reports in the published literature, including current management guidelines.

How does this study impact current management guidelines? The MCID calculations using the validated anchor-based estimate reported here support most of the recommendations of current management guidelines. The finding that intranasal corticosteroid with intranasal antihistamine in the same device was more effective than either monotherapy alone should be carefully reviewed for future guidance documents. In addition, we believe that the approach used in this article currently represents the only reasonable method to determine an MCID for allergic rhinitis studies and should supersede the method and consequent findings of the AHRQ report.

BACKGROUND: In 2013, the Agency for Healthcare Research and Quality (AHRQ) recommended that allergic rhinitis (AR) studies calculate a minimal clinically important difference (MCID) based on an estimated threshold equal to 30% of the maximum total nasal symptom score. Applying this threshold, their data showed no differences between well-established treatments, and a subsequent analysis using prescribing information found no differences between active treatments and placebo controls.

OBJECTIVE: The objective of this study was to demonstrate the application of an evidence-based model to determine MCIDs for
AR studies, with an absolute value for an anchor-based threshold and validated methods for calculating distribution-based thresholds.

METHODS: Using the same studies as the AHRQ report, anchor- and distribution-based MCID thresholds were determined for 3 clinical comparisons identified by the AHRQ: (1) oral antihistamine + intranasal corticosteroid (INCS) versus INCS, (2) montelukast versus INCS, and (3) intranasal antihistamine + INCS in a single device versus the monotherapies. The outcomes were compared with those reported using the AHRQ threshold.

RESULTS: No treatment comparison met the AHRQ-de determined threshold; all treatments were determined to be equivalent.

CONCLUSIONS: The evidence-based threshold for MCID determination for AR studies should supersede the threshold recommended in the AHRQ report.

Evidence-based medicine integrates research outcomes, clinical expertise, and patient expectations to optimize clinical decision making during treatment. A key pillar of the evidence-based approach is the concept that, to be considered effective, a therapy should provide both statistically significant and clinically meaningful differences over a placebo and/or active comparators. What is clinically meaningful can be estimated through determination of a minimal clinically important difference (MCID), which is defined as the minimal amount of a treatment effect (or change) that is important to the patient.1,2 How to measure this in a manner that incorporates the patient’s perspective yet allows for appropriate comparison of different treatments is subject to discourse.

Multiple evidence-based methods for determining an MCID have been described, with most falling into 2 classes: anchor-based and distribution-based approaches. Both can be used to determine the magnitude of a clinically relevant treatment effect size from a population perspective that is, quantitatively, based on treatment group means.1,2,5

As named, the anchor-based approach links a change in a desired outcome measure to a “meaningful” external anchor that reflects the patient’s perspective, such as the global rating of change score (GRCS) by which patients rate their impression of treatment.1,2,6,7 For example, the patient might be asked to finish the statement, “Since starting therapy my symptoms are,” using an ordinal scale from −7 (very much worse) to 0 (no change) to +7 (very much better).

Distribution-based approaches assess statistically significant changes in the desired outcome measure in relation to the probability of change occurring by chance. For example, a clinically meaningful effect might be defined as a change above an arbitrary multiple of the sample standard deviation (SD) for the measure at baseline.1,2,6 Because distribution-based methods are sample specific, MCID scores can be determined by statistical analysis alone, even when a change from baseline is difficult to detect (eg, in studies with large sample sizes and variances).2 However, unlike anchor-based approaches, distribution-based calculations are not necessarily linked to any patient perspective of a clinically meaningful response. Consequently, anchor-based MCIDs are generally considered more robust.1,2,7,8

Determining an MCID in allergic rhinitis studies

How MCID comparisons apply to clinical decision making varies by disease state.1,2,7 For some, including allergic rhinitis (AR), how to calculate the MCID remains a point of discussion. To date few articles have addressed this issue for AR, and those that have—including guidances from government health care agencies in the European Union and the United States—suggest widely different approaches (see Appendix E1 available in this article’s Online Repository at www.jaci-inpractice.org).1,3,8-12

For the patient, AR is a disease characterized by annoying symptoms, and, reflecting this, the most commonly used scale to
Assess efficacy in AR is the total nasal symptom score (TNSS), which is typically the sum of the individual symptom scores for nasal itching, rhinorrhea, sneezing, and nasal congestion. In most clinical trials, each of these symptoms is rated on a scale from 0 to 3 (0 = none, 1 = mild, 2 = moderate, 3 = severe) twice daily. Both morning + evening (AM+PM) sums and averages have been reported in the literature, so the TNSS can range from 0 to 12 or 0 to 24 according to the study design. Less frequently, the TNSS may be rated on a 0- to 4-point scale or, in some studies, symptoms may be rated on a visual analog scale (VAS) from 0 to 100 mm, resulting in a TNSS range of 0-400 mm. Nonetheless, the TNSS is the most accepted primary efficacy variable that is rated for drug approval in AR in the USA, so, when considering clinical relevance, it is appropriate to assess MCID in relation to the TNSS.

The different methods proposed for determining MCIDs in AR studies (Appendix E1, available in this article’s Online Repository at www.jaci-inpractice.org) underscore the need to better understand how to evaluate what is a meaningful change to the patient in a manner that is both evidence based and able to compare treatment means. A recent attempt to do this was published by the Agency for Healthcare Research and Quality (AHRQ) in the USA in 2013. The agency sought to develop an approach for determining an MCID threshold to evaluate drug classes used to treat seasonal AR (SAR), including oral and nasal antihistamines and decongestants; intranasal corticosteroids (INCS); the mast cell stabilizer, intranasal cromolyn sodium; the antihistamines and decongestants; intranasal corticosteroids (INCS); the mast cell stabilizer, intranasal cromolyn sodium; the anticholinergic, intranasal ipratropium; the oral leukotriene receptor antagonist (LTRA), montelukast; and nasal saline. The evaluation was conducted by a panel convened by the Blue Cross and Blue Shield Association Technology Evaluation Center, using the TNSS as the measure of effectiveness based on meta-analyses of studies evaluating single treatments and combinations of treatments. According to that panel, “anchor-based MCIDs have not been defined for rhinitis symptom scales.” Thus, a subpanel of 3 recommended post hoc an MCID equal to 30% of the maximum TNSS (eg, ±3.6 points on a 12-point scale) as a useful threshold. Treatment differences that failed this threshold would indicate equivalence. The panel acknowledged that validation of this definition of MCID using an anchor-based approach would be desirable.

Applying this threshold, the AHRQ panel could not demonstrate any differences in effectiveness between the various therapeutic classes, which they mostly attributed to insufficient evidence to support the superiority of one treatment over another. Although the lack of good comparative data for some of the comparisons certainly contributed to the outcomes, of greater concern is that the AHRQ method was flawed in 2 important ways.

The first arises from using the maximum possible score of a bounded outcome to determine the threshold. This method creates problems around the lower bound. In the case of the AHRQ threshold, using the fixed number of ±3.6 points (on a 12-point scale) based on 30% of the maximum TNSS could ultimately negate milder levels of AR from being clinically relevant. Specifically, any study population with an average baseline score of 3.5 or less could not attain the MCID threshold of 3.6, even when treatment eliminated all symptoms in every patient (ie, posttreatment value = 0).

Second, although the 30% criterion could be relevant for an individual patient response, there was no indication of how it could be applied to a comparison of differences in population means. Applying the panel members’ opinion of patient response to population-based differences in treatment means from clinical trials resulted in overestimation of the MCID threshold, which can be demonstrated by comparing the conclusions based on the application of the AHRQ threshold directly with those based on a threshold derived from an earlier published method authored by Barnes et al.

Barnes et al developed and tested anchor- and distribution-based models to determine MCID thresholds for AR treatments, using pooled data from 9 randomized, placebo-controlled studies (n = 204 subjects with mild-to-moderate AR). These evidence-based models provide clinicians with an absolute value for an anchor-based threshold and validated methods for calculating distribution-based thresholds.

The anchor-based MCID threshold value for a TNSS scale of 0-12 points was 0.23 or 0.28 points depending on whether regression or meta-analytical methods, respectively, were applied, and was derived from calculations establishing a relationship between GRCS (as a direct anchor) and TNSS. For a TNSS scale of 0-24 points, the comparable MCID thresholds would be 0.46 points (by regression analysis) or 0.56 points (by meta-analysis), and for a 400 mm VAS, the upper threshold was 9.33 mm.

The distribution-based model applied commonly used statistical methods for determining effect size by measuring the distance between 2 treatment means in relation to the baseline SD of the sample (eg, Hedge’s g, Cohen’s d). Validity analyses for all the methods tested were described in detail. For more information about the calculations, the reader is referred to Appendix E2, available in this article’s Online Repository at www.jaci-inpractice.org.

Brixner et al applied the higher, more conservative anchor-based threshold of 0.28 points to TNSS data reported in the approved prescribing information for intranasal azelastine hydrochloride (AZE), ciclesonide, fluticasone furoate, and the combination of azelastine and fluticasone propionate (FP) in a single device (MP-AzeFlu), and then compared those MCID outcomes with the 30% threshold recommended by the AHRQ panel. All 4 products achieved the threshold of a “clinically meaningful” change compared with placebo using the anchor-based estimate; in contrast, none showed any clinical benefit over placebo based on the AHRQ 30% threshold. These observations suggest that applying the Barnes models to the dataset used in the AHRQ report would yield different results and would support the Barnes et al recommendations for how to best determine an MCID in AR studies.

**METHODS**

In a new series of assessments, anchor- and distribution-based MCID thresholds using the estimates and methods described by Barnes et al were compared with the AHRQ report (2013) recommended threshold of 30% of the TNSS maximum for 3 treatment-related clinical questions evaluated in the AHRQ report:

1. Is there any clinical benefit to adding an oral antihistamine (OAH) to an INCS?
2. How does the LTRA, montelukast, compare with INCSs in terms of clinical benefit?
3. Is there any clinical benefit to adding an intranasal antihistamine (INAH) to an INCS in a single device?

The questions were chosen based on outcomes in the AHRQ report suggesting that the treatments were equivalent in terms of...
efficacy for nasal symptoms, with a high degree of confidence—findings that are in opposition to other published data, including national and international guideline recommendations, and to the clinical experience of the authors of this article.

The clinical studies described in the AHRQ report provided the initial dataset for analysis. The inclusion criteria were studies of SAR, at least 2 weeks in duration. All subjects had a minimum of a 2-year clinical history of SAR of mild-to-moderate severity, with positive skin prick test results in the year before study, but otherwise were in good health.

Studies published subsequent to the AHRQ report were searched using a similar strategy to that described by the AHRQ panel: MEDLINE, Embase, and the Cochrane Library were searched for articles in English reported between July 18, 2012, and September 8, 2015. Only search terms relevant to the 3 queries were searched for articles in English reported between July 18, 2012, and September 8, 2015. Only search terms relevant to the 3 queries were searched for articles in English reported between July 18, 2012, and September 8, 2015. Only search terms relevant to the 3 queries were searched for articles in English reported between July 18, 2012, and September 8, 2015. Only search terms relevant to the 3 queries were searched for articles in English reported between July 18, 2012, and September 8, 2015.

RESULTS

Only studies from the AHRQ report (2013) were used in the dataset; the extended search found no additional studies meeting the AHRQ criteria for inclusion. The studies are described in Appendix E4 (available in this article’s Online Repository at www.jaci-inpractice.org).

Tables I-III summarize the MCID outcomes for the different approaches by query. Overall, none of the treatment comparisons in the studies assessed in the AHRQ met the 30% threshold recommended by their panel to define a minimal clinically important difference between treatments. Thus, all treatments were determined to be equivalent for each of the 3 queries.

When using the anchor-based threshold or the respective Hedge’s $g$ or Cohen’s $d$ was performed to further assess the results of the threshold findings. The calculation, which is study specific and is based on both the difference in treatment means and the weighted average baseline SD, is described in full in Appendix E3 (available in this article’s Online Repository at www.jaci-inpractice.org). The MCID threshold for Hedge’s $g$ (or Cohen’s $d$) is $\pm 0.2$.

DISCUSSION

To date, there is one published report describing validated evidence-based methods for determining an MCID in AR studies, including both anchor-based and distribution-based approaches—the paper authored by Barnes et al published in 2010. Their anchor-based thresholds can be directly applied to AR clinical studies with appropriate data, and the relatively simple distribution-based calculation (see Appendix E2, available in this article’s Online Repository at www.jaci-inpractice.org) can be used to support borderline anchor-based outcomes.

Application of these methods and thresholds to 3 of the same queries evaluated in the AHRQ report and using their database yielded different outcomes from those reported by the AHRQ panel. Specifically, in terms of reaching an MCID for patient symptom relief, we found that (1) OAH + INCS $\sim$ INCS, (2) INCS $>$ montelukast, and (3) MP-AzeFlu or AZE $>$ Aze or FP. These outcomes could be expected from other reviews and meta-analyses in the literature and from physicians’ clinical experience. As such, the methods and estimates reported by Barnes et al are recommended to determine an MCID in AR studies.

At this time, other than the approach described in the Barnes paper, there are no other appropriate and validated methods to...
### TABLE I. Query 1: Is there any clinical benefit for adding an oral antihistamine to an intranasal corticosteroid (INCS)?

<table>
<thead>
<tr>
<th>Study</th>
<th>Tx</th>
<th>Difference between Tx in TNSS change from BL @ 2 wk</th>
<th>Hedge’s <em>g</em></th>
<th>MCID threshold met?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anolik et al, 2008</td>
<td>MOM+LOR vs MOM</td>
<td>−0.3</td>
<td>−0.13</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Barnes et al, 2006</td>
<td>FP+LCET vs FP</td>
<td>−0.11</td>
<td>NA</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ratner et al, 1998</td>
<td>FP+LOR vs FP</td>
<td>+1.0†</td>
<td>+0.02</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

**MCID threshold met?**

- **AHRQ**: Agency for Healthcare Research and Quality
- **Anchor based**: magnitude of Hedge’s *g* ≥ 0.20 SD
- **Distribution based**: magnitude of Hedge’s *g* ≥ 0.20 SD
- **Comments**: Reported data lack precision; assay sensitivity is confirmed by Tx differences for monotherapy arms vs P

### TABLE II. Query 2: How does montelukast compare with an intranasal corticosteroid in terms of clinical benefit?

<table>
<thead>
<tr>
<th>Study</th>
<th>Tx</th>
<th>Difference between Tx in TNSS change from BL @ 2 wk</th>
<th>Hedge’s <em>g</em></th>
<th>MCID threshold met?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lu et al, 2009</td>
<td>Study 1: BDP vs MON</td>
<td>−0.34</td>
<td>NA</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Yes</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Martin et al, 2006</td>
<td>FP vs MON</td>
<td>−33.6 †</td>
<td>−0.63</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nathan et al, 2005</td>
<td>FP vs MON</td>
<td>−26.1 †</td>
<td>−0.33</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ratner et al, 2003</td>
<td>FP vs MON</td>
<td>−36.3 †</td>
<td>−0.72</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pullerits et al, 2002</td>
<td>FP vs MON</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

**MCID threshold met?**

- **AHRQ**: Agency for Healthcare Research and Quality
- **Anchor based**: magnitude of Hedge’s *g* ≥ 0.20 SD
- **Distribution based**: magnitude of Hedge’s *g* ≥ 0.20 SD

### Comments

- **AHRQ**: Agency for Healthcare Research and Quality
- **Anchor based**: magnitude of Hedge’s *g* ≥ 0.20 SD
- **Distribution based**: magnitude of Hedge’s *g* ≥ 0.20 SD

**Comments**: Reported data lack precision; assay sensitivity is confirmed by Tx differences for monotherapy arms vs P

**Physician-evaluated Sx scores**

**MCID threshold met?**

- **AHRQ**: Agency for Healthcare Research and Quality
- **Anchor based**: magnitude of Hedge’s *g* ≥ 0.20 SD
- **Distribution based**: magnitude of Hedge’s *g* ≥ 0.20 SD

**MCID threshold met?**

- **AHRQ**: Agency for Healthcare Research and Quality
- **Anchor based**: magnitude of Hedge’s *g* ≥ 0.20 SD
- **Distribution based**: magnitude of Hedge’s *g* ≥ 0.20 SD

- **Changes** in symptoms of first drug/drug combination when compared with second drug/drug combination and third drug/drug combination.
- **AHRQ** panel threshold: 30% difference in maximum TNSS change from BL (ie, ±3.6 on a 0-12 scale, 7.2 on a 0-24 scale, or 120 mm on a 400 mm VAS).8
- **Anchor-based estimate**: ±0.28 for a scale of 0-12, ±0.56 for a scale of 0-24 (indicated by †) and ±9.33 mm for a 400 mm VAS.8
- **Distribution-based threshold**: magnitude of Hedge’s *g* ≥ 0.20 SD (see the text and Appendix E1, available in this article’s Online Repository at www.jaci-impractice.org).

**MCID threshold met?**

- **AHRQ**: Agency for Healthcare Research and Quality
- **Anchor based**: magnitude of Hedge’s *g* ≥ 0.20 SD
- **Distribution based**: magnitude of Hedge’s *g* ≥ 0.20 SD

**MCID threshold met?**

- **AHRQ**: Agency for Healthcare Research and Quality
- **Anchor based**: magnitude of Hedge’s *g* ≥ 0.20 SD
- **Distribution based**: magnitude of Hedge’s *g* ≥ 0.20 SD

**MCID threshold met?**

- **AHRQ**: Agency for Healthcare Research and Quality
- **Anchor based**: magnitude of Hedge’s *g* ≥ 0.20 SD
- **Distribution based**: magnitude of Hedge’s *g* ≥ 0.20 SD

**MCID threshold met?**

- **AHRQ**: Agency for Healthcare Research and Quality
- **Anchor based**: magnitude of Hedge’s *g* ≥ 0.20 SD
- **Distribution based**: magnitude of Hedge’s *g* ≥ 0.20 SD

**MCID threshold met?**

- **AHRQ**: Agency for Healthcare Research and Quality
- **Anchor based**: magnitude of Hedge’s *g* ≥ 0.20 SD
- **Distribution based**: magnitude of Hedge’s *g* ≥ 0.20 SD

**MCID threshold met?**

- **AHRQ**: Agency for Healthcare Research and Quality
- **Anchor based**: magnitude of Hedge’s *g* ≥ 0.20 SD
- **Distribution based**: magnitude of Hedge’s *g* ≥ 0.20 SD

**MCID threshold met?**

- **AHRQ**: Agency for Healthcare Research and Quality
- **Anchor based**: magnitude of Hedge’s *g* ≥ 0.20 SD
- **Distribution based**: magnitude of Hedge’s *g* ≥ 0.20 SD

**MCID threshold met?**

- **AHRQ**: Agency for Healthcare Research and Quality
- **Anchor based**: magnitude of Hedge’s *g* ≥ 0.20 SD
- **Distribution based**: magnitude of Hedge’s *g* ≥ 0.20 SD

- **Changes** in symptoms of first drug/drug combination when compared with second drug/drug combination and third drug/drug combination.
- **AHRQ** panel threshold: 30% difference in maximum TNSS change from BL (ie, ±3.6 on a 0-12 scale, 7.2 on a 0-24 scale, or 120 mm on a 400 mm VAS).8
- **Anchor-based estimate**: ±0.28 for a scale of 0-12, ±0.56 for a scale of 0-24 (indicated by †) and ±9.33 mm for a 400 mm VAS.8
- **Distribution-based threshold**: magnitude of Hedge’s *g* ≥ 0.20 SD (see the text and Appendix E1, available in this article’s Online Repository at www.jaci-impractice.org).
responder analyses to evaluate the proportion of patients with 50% or more reduction in a symptom score may, for individual studies, add a measure of clinical relevance to statistically significant improvements in symptom scores.9,22,23 However, responder analysis has not yet been broadly applied, and would require validation to be used as a stand-alone global effect measure. The World Allergy Organization recommended using a relative clinical impact score, which evaluates the clinical effect of an active treatment relative to the effect of a matched placebo.10,36 Although this has been used to determine clinical benefit for a meta-analysis of sublingual allergen immunotherapy (SLIT) and pharmacotherapy in SAR,36 it is an indirect method that needs validation. Comparable baseline scores are necessary for between-treatment comparisons, and paradoxical effects may arise when comparing prophylactic treatments such as SLIT with in-season pharmacotherapy as well as from studies of intranasal medications in which the intranasal placebos could, depending on the amount of fluid volume, possibly act as a nasal rinse. These considerations serve to underscore the fact that, for now, the Barnes approach is the only reasonable method to determine an MCID for AR studies.

Further evaluation of the Barnes approach is needed—particularly in terms of expanding its application. Questions remain as to whether the methods could be applied to longer term studies of perennial AR and chronic non-AR as well as to other types of studies such as prophylaxis, with treatment started before a pollen season. Application of the approach to risk of AHRQ panel threshold: 30% difference in maximum TNSS change from BL (ie, ±3.6 on a 0-12 scale, 7.2 on a 0-24 scale, or 120 mm on a 400 mm VAS).† Distribution-based threshold: magnitude of Hedge’s g ≥ 0.2 SD (see the text and Appendix E1, available in this article’s Online Repository at www.jaci-inpractice.org).

### TABLE III. Query 3: Is there any clinical benefit to adding an intranasal antihistamine to an intranasal corticosteroid (INCS) in a single device?

<table>
<thead>
<tr>
<th>Study</th>
<th>Tx</th>
<th>Difference between Tx in TNSS change from BL @ 2 wk*</th>
<th>Hedge’s g*</th>
<th>AHRQ†</th>
<th>Anchor based‡</th>
<th>Distribution based‡</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hampel et al, 2010</td>
<td>MP29-02, aka</td>
<td>MP-AzeFlu vs FP</td>
<td>-1.47†</td>
<td>0.45</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AZE</td>
<td>-2.06†</td>
<td>0.61</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Carr et al, 2012 Study</td>
<td>MP29-02, aka</td>
<td>MP-AzeFlu vs FP</td>
<td>-1.0†</td>
<td>-0.32</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AZE</td>
<td>-1.0†</td>
<td>-0.31</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Carr et al, 2012 Study</td>
<td>MP29-02, aka</td>
<td>MP-AzeFlu vs FP</td>
<td>-0.6†</td>
<td>-0.25</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AZE</td>
<td>-0.7†</td>
<td>-0.29</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Carr et al, 2012 Study</td>
<td>MP29-02, aka</td>
<td>MP-AzeFlu vs FP</td>
<td>-0.9†</td>
<td>-0.29</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AZE</td>
<td>-1.4†</td>
<td>-0.43</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

*Change (greater (+) or less (−)) in reduction in symptoms of first drug/drug combination when compared with second drug/drug combination and third drug/drug combination.
†AHRQ panel threshold: 30% difference in maximum TNSS change from BL (ie, ±3.6 on a 0-12 scale, 7.2 on a 0-24 scale, or 120 mm on a 400 mm VAS).
‡Anchor-based estimate: ±0.28 for a scale of 0-12, ±0.56 for a scale of 0-24 (indicated by ††), and ±9.33 mm for a 400 mm VAS.
§Distribution-based threshold: magnitude of Hedge’s g ≥ 0.2 SD (see the text and Appendix E1, available in this article’s Online Repository at www.jaci-inpractice.org).

In conclusion, our recommendation is that the method of Barnes et al1 for determining an MCID for AR studies and the conclusions based on that method should supersede the method and consequent findings of the AHRQ report.8
Acknowledgments
The authors would like to acknowledge both the editorial and technical support of Judith Rosen Farrar, PhD, FAAAAI, and the statistical expertise of Ulrich Munzel, PhD.

REFERENCES
1. Barnes ML, Vaidyanathan PA, Williamson PA, Lipworth BJ. The minimal clinically important difference in allergic rhinitis. Clin Exp Allergy 2010;36:676-84.
5. Brixner D, Meltzer EO, Morland K, Carroll CA, Lipworth BJ. The importance of anchor based minimal clinically important difference (MCID) to health technology assessment of established intranasal allergic rhinitis treatments. Poster presented at: Annual Meeting of the International Society for Pharma-neconomics and Outcomes Research; May 16-20, 2015: Philadelphia, PA.
### ONLINE REPOSITORY

#### APPENDIX E1. Some proposed methods for determining MCID thresholds for allergic rhinitis (AR) treatments

<table>
<thead>
<tr>
<th>Method</th>
<th>How used and/or developed</th>
<th>Rating</th>
<th>Comments</th>
<th>An appropriate global effect measure for AR?</th>
</tr>
</thead>
<tbody>
<tr>
<td>EU/CHMP Guideline (2004)(^{E1})</td>
<td>Suggested in guidance as a means to evaluate clinical benefit of medicinal products for allergic rhinoconjunctivitis</td>
<td>Responder analyses based on the proportion of patients with ( \geq 50% ) reduction in a Sx score</td>
<td>A change in score ( \geq 3 ) has generally been considered as a clinically relevant improvement of Sx based on reductions from BL in rTSS4 or TNSS of 20%-35% for P and 40%-50% for active Tx</td>
<td>No, requires access to database(s) for studies, and has other limitations noted(^{E1})</td>
</tr>
<tr>
<td>WAO relative clinical impact(^{E2,E3})</td>
<td>Applied to immunotherapy</td>
<td>Relates differences in active Tx effects to placebo score</td>
<td>Appropriate for systemic prophylaxis when Tx is initiated before allergy season and BL severity is not considered</td>
<td>No</td>
</tr>
<tr>
<td>Barnes et al (2010)(^{E4})</td>
<td>To determine MCIDs for AR Tx in clinical studies based on TNSS</td>
<td>A validated, quantitative anchor-based approach(^{E4})</td>
<td>The GRCS was used as a direct anchor, yielding MCID thresholds of 0.23-0.28 points for a TNSS scale of 0-12</td>
<td>Yes</td>
</tr>
<tr>
<td>Cohen’s (d) and Hedge’s (g)(^{E4,E5})</td>
<td>A statistical calculation that relates the change in an outcome of interest to some measure of its normal variability, for example, the sample SD of the TNSS at BL</td>
<td>A validated, quantitative, distribution-based approach</td>
<td>Cohen’s (d) recommends any change ( \geq 0.2 ) times the BL SD as clinically meaningful Hedge’s (g) uses a weighted average of the BL SDs, which is the same approach to approximate the SD as the ( t)-test(^{E6})</td>
<td>Easy to calculate when the information regarding variability is present</td>
</tr>
</tbody>
</table>

(continued)
APPENDIX E1. (Continued)

<table>
<thead>
<tr>
<th>Method</th>
<th>How used and/or developed</th>
<th>Rating</th>
<th>Comments</th>
<th>An appropriate global effect measure for AR?</th>
</tr>
</thead>
<tbody>
<tr>
<td>AHRQ threshold(^4) (AHRQ, 2013)</td>
<td>To evaluate SAR treatments based on TNSS</td>
<td>Arbitrary panel-based recommendation of 30% change in TNSS based on the total scale range (eg, (\pm 3.6) points based on a 12-point TNSS scale)</td>
<td>Used to assess meta-analyses of single Tx and combinations of Tx, but could not demonstrate clinically meaningful benefit between different Tx, or even between active Tx and placebo control</td>
<td>No</td>
</tr>
</tbody>
</table>

**LIMITATIONS:**
- A subjective, nonvalidated approach
- Attempted to apply a patient-based perspective of MCID to a population mean difference, resulting in a gross overestimation of the MCID threshold

BL, Baseline; GRCS, global rating of change score; MCID, minimal clinically important difference; P, placebo; rTSS\(^4\), rhinitis total symptom score 4 (similar to TNSS); SAR, seasonal allergic rhinitis; SD, standard deviation; Ss, symptoms; TNSS, total nasal symptom score; Ts, treatment(s).

APPENDIX E2: DISTRIBUTION CALCULATION OF MCID THRESHOLD BASED ON BARNES ET AL\(^4\)

The distribution-based calculation described by Barnes et al\(^4\) used to evaluate MCID thresholds for AR treatments was Hedge’s \(g\), a modification of Cohen’s \(d\). The latter is a commonly used method to determine effect size by measuring the distance between 2 means in relation to the baseline standard deviation of the samples.\(^4\),\(^6\) For treatment effects in clinical trials in AR, the calculation would be: \(^4\),\(^6\),\(^8\)-\(^10\)

\[
d = \frac{(\text{Treatment Mean 1} - \text{Treatment Mean 2})}{(\text{Baseline Standard Deviation for the pooled samples})}
\]

In general, Cohen’s \(d\) recommends as clinically meaningful any treatment difference \(\geq 0.2\) times the baseline standard deviation, with effect sizes of 0.2, 0.5, and 0.8 standard deviations suggested as small, moderate, and large effects, respectively.\(^4\),\(^6\)

Hedge’s \(g\) uses a weighted average in the denominator of the calculation, which may be preferable to Cohen’s \(d\) because it is the same approach to approximate standard deviation as used in the common \(t\)-test.\(^4\),\(^6\),\(^8\)-\(^10\)

\[
g = \frac{(\text{Treatment Mean 1} - \text{Treatment Mean 2})}{(\text{weighted average of the Baseline Standard Deviations for both samples})}
\]

As for Cohen’s \(d\), a treatment difference \(\geq 0.2\) times the baseline standard deviation is considered clinically meaningful for Hedge’s \(g\).\(^4\),\(^6\),\(^8\)-\(^10\)

Applying the Hedge’s \(g\) calculation to the 9 pooled studies described in their paper, Barnes et al\(^4\) were able to define a common estimate of the baseline standard deviation and multiplying that result by 0.2 (for a clinically meaningful change) resulted in an MCID threshold of 0.59 for the TNSS on a 12-point scale (1.18 units on a 24-point scale) for those studies.\(^4\),\(^6\),\(^8\)-\(^10\) It is important to recognize that because these calculations are sample specific, the threshold determined (eg, 0.59 for the TNSS on a 12-point scale for the Barnes et al pooled studies\(^4\)) is also specific to that sample, so that it is better to recalculate \(g\) in each study anew or to use a general estimate such as the anchor-based approach.\(^4\),\(^6\),\(^8\)-\(^10\)

As an example, the reader is directed to the data from Study MP4004 reported in the paper by Carr et al\(^1\) (see Tables 2 and 3c). The baseline standard deviation for the MP29-02 treatment group was 3.3 with \(N = 193\); the baseline standard deviation for the fluticasone propionate (FP) group was 2.9 with \(N = 189\). Therefore, the weighted average of the baseline standard deviations for these 2 treatment groups is 3.10 using the following calculation:

*Weighted average of BL SDs* = \(\frac{3.3 \times 192 + 2.9 \times 188}{192 + 188}\) = 3.10.

The treatment means for the 2 groups at 2 weeks were MP-AzeFlu (aka MP29-02), 12.6; FP, 13.6. The difference in the treatment group means (MP-AzeFlu – FP) = \(-1.0\), so the calculation for Hedge’s \(g\) is as follows:

\[
\text{Hedge’s } g = \frac{-1.0}{3.1} = -0.32.
\]

The absolute value (0.32) exceeds the MCID threshold for Hedge’s \(g\) of 0.20. Therefore, the difference between the 2 treatment groups is considered to be at least minimally clinically relevant, with MP-AzeFlu > FP.
APPENDIX E3: DESCRIPTION OF THE SEARCH STRATEGY TO EXTEND THE LITERATURE SEARCH FOR QUERIES 1, 2, AND 3

The dataset for analysis was extended to current (as of September 8, 2015) using a modification of the search strategy described in the AHRQ report and the same electronic databases—MEDLINE, EMBASE, and the Cochrane Library. The following limits were applied: English, human subjects, dates 2012 (only studies published after July 18, 2012, considered) to current (September 8, 2015).

The modification involved adding the term *human subjects* as a limit and simplifying some of the search terms for a more direct approach appropriate to the scope of this article (eg, clinical trial replaced placebo-controlled trial + controlled trial + randomized controlled trial + case cohort study + observational trial + cross-sectional study). The search terms were applied as MeSH descriptors, headers, and/or simple search terms according to the structure of each database, and then combined as shown in Appendix E3 Figure 1. Only search terms relevant to the 3 queries were included.

Any citations were reviewed for inclusion criteria as noted in the text—seasonal allergic rhinitis, minimum of 2-week trial duration, mild-to-moderate disease severity, symptoms scored by TNSS, GRCS, or TSS4, and direct comparisons between treatments as indicated by each query.

Overall, for the individual search terms (with limits applied), MEDLINE yielded 4714 records, EMBASE yielded 6177 records, and the Cochrane Library yielded 17 records. The combined totals from each database and the number of records specific to each query are shown in Appendix E2 Figure 1. After combining the terms for the individual queries, only 3 citations were determined to be appropriate for review under query 3: a poster presentation and 2 long-term studies of perennial and chronic rhinitis. None met the inclusion criteria for the studies described by the AHRQ report.

APPENDIX E4: STUDY DESCRIPTIONS

Studies of treatments for seasonal allergic rhinitis (SAR) that were used in the AHRQ evaluation of MCIDs are included as described in the text. The TNSS presented is an average of AM/PM scores reported by the subject unless otherwise noted.

Patient inclusion criteria:
- A 2-year clinical history of SAR unless otherwise indicated;
- good health; no clinically significant disease other than SAR;
- positive skin prick test results in past year.

---

**APPENDIX E3 Figure 1.** Search strategy adapted from AHRQ Report. The number of records obtained is shown for each of 5 searches, followed by the number of records remaining after combining the appropriate searches for each of 3 queries. The number in parentheses indicates how many citations were reviewed more closely for possible inclusion in the study table.
### APPENDIX E4 TABLE 1. Study descriptions

<table>
<thead>
<tr>
<th>Study</th>
<th>Study design and length of active Tx</th>
<th>TNSS scale maximum</th>
<th>Tx (n)</th>
<th>TNSS ± SE/SD at BL</th>
<th>Absolute change in TNSS ± SE/SD (if available) at 2 wk</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Query 1. Is there any clinical benefit to adding an oral antihistamine to an intranasal corticosteroid (INCS)?</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anolik et al, 2008</td>
<td>R, DB, PG, PC; 15 d</td>
<td></td>
<td>12</td>
<td>7.9 ± 2.0</td>
<td>-3.0 ± 2.0</td>
<td>Values presented are means ± SD</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7.8 ± 2.5</td>
<td>-2.7 ± 2.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7.9 ± 2.2</td>
<td>-1.9 ± 2.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8.0 ± 2.2</td>
<td>-1.4 ± 2.2</td>
<td></td>
</tr>
<tr>
<td>Barnes et al, 2006</td>
<td>R, DB, PC, crossover; 2 wk each phase</td>
<td></td>
<td>12</td>
<td>4.56 ± 2.58</td>
<td>-2.13</td>
<td>Values presented are means ± SE where available</td>
</tr>
<tr>
<td>Ratner et al, 1998</td>
<td>R, DD, DB, PG, PC; 2 wk</td>
<td>400 (VAS)</td>
<td></td>
<td>304.9 ± 4.7</td>
<td>-186.0 ± 9.4</td>
<td>Physician-evaluated Sx scores</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>304.9 ± 4.6</td>
<td>-187.0 ± 8.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>313.3 ± 4.0</td>
<td>-102.0 ± 9.9</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>302.4 ± 4.2</td>
<td>-102.0 ± 8.8</td>
<td></td>
</tr>
<tr>
<td><strong>Query 2. How does montelukast compare with an intranasal corticosteroid in terms of clinical benefit?</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lu et al, 2009</td>
<td>Study 1*: R, DB, MC, PG, PC; 2 wk</td>
<td>3</td>
<td>Study 1:</td>
<td>2.06</td>
<td>-0.36</td>
<td>DT NSS = average of 4 nasal Sx scores, each on a 0-3 scale; no SD or SE reported</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MON 10 mg QD (111)</td>
<td>2.11</td>
<td>-0.53</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LOR 10 mg QD (115)</td>
<td>2.04</td>
<td>-0.54</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MON + LOR QD (174)</td>
<td>2.03</td>
<td>-0.70</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>BDP 200 µg BID (172)</td>
<td>2.04</td>
<td>-0.22</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>P (56)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Martin et al, 2006</td>
<td>R, DB, DD, PG, MC; 2 wk</td>
<td>400 (VAS)</td>
<td></td>
<td>298.2 ± 2.8</td>
<td>-130.2 ± 4.7</td>
<td>Values presented are means ± SE</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>301.5 ± 2.8</td>
<td>-96.6 ± 4.7</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>DT TNSS only</td>
</tr>
<tr>
<td>Nathan et al, 2005</td>
<td>R, DB, PG, PC, MC; 4 wk</td>
<td>400 (VAS)</td>
<td></td>
<td>260.7 ± 4.6</td>
<td>-99.1 ± 5.8</td>
<td>Patients also had asthma and were using FP and/or salmeterol BID</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>269.1 ± 4.7</td>
<td>-73.0 ± 1.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>260.5 ± 4.5</td>
<td>-60.7 ± 5.8</td>
<td>DT TNSS for wk 1-2 only.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Values presented are means ± SE</td>
</tr>
<tr>
<td>Ratner et al, 2003</td>
<td>R, DB, DD, PG, MC; 15 d</td>
<td>400 (VAS)</td>
<td></td>
<td>296.2 ± 2.7</td>
<td>-130.3 ± 4.7</td>
<td>Daytime TNSS only. Values presented are means ± SE</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>298.9 ± 2.7</td>
<td>-94.0 ± 4.7</td>
<td></td>
</tr>
<tr>
<td>Pullerits et al, 2002</td>
<td>R, DB, DD, PG, PC; 8 wk (allergy season)</td>
<td>16 (used scale of 0-4)</td>
<td></td>
<td>TNSS day:</td>
<td>1.5 ± 1.4</td>
<td>Values presented at BL are means</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.9 ± 2.1</td>
<td>1.7 ± 0.7</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.9 ± 1.5</td>
<td>2.6 ± 0.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.4 ± 2.3</td>
<td>2.1 ± 0.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TNSS night:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.9 ± 1.2</td>
<td>3.5 ± 0.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>&lt;20 subjects per Tx arm.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Different type of comparison as Tx began at start of 8-wk pollen season so that Sx stability and not change in Sx was assessed</td>
</tr>
</tbody>
</table>
**Query 3:** Is there any clinical benefit to adding an intranasal antihistamine to an INCS in a single device?

| Study | Design | DB, PC, PG, R, MC; | Duration | FP | AZE | MP29-02, aka MP-AzeFlu | P | All | **AM** | **PM** | | | | | | **Ty** | **SE** | **SD** | **Change** | Note |
|-------|--------|-------------------|----------|----|-----|-----------------------|----|-----|-------|-------|----------|--------|--------|-----|--------|------|--------|------|
| Hampel et al, 2010; Meltzer et al, 2013 | DB, PC, PG, R, MC; 2 wk | 24 | FP 100 µg (151) | AZE 274 µg (152) | MP29-02, aka MP-AzeFlu (153) | P (151) | All BID | 18.3 | 18.1 | 18.8 | 18.7 | −3.84 | ± SE | ± SD; the changes are mean TNSS averaged over wk 1-2 |
| Carr et al, 2012 Study MP4004 | DB, PC, PG, R, MC; 2 wk | 24 | FP 100 µg (189) | AZE 274 µg (194) | MP29-02, aka MP-AzeFlu (193) | P (200) | All BID | 18.6 | 18.5 | 18.2 | 18.2 | −5.0 | ± SE | ± SD |
| Carr et al, 2012 Study MP-4006 | DB, PC, PG, R, MC; 2 wk | 24 | FP 100 µg (450) | AZE 274 µg (445) | MP29-02, aka MP-AzeFlu (448) | P (448) | All BID | 19.4 | 19.5 | 19.4 | 19.5 | −5.1 | ± SE | ± SD |
| Carr et al, 2012 Study MP-4002 | DB, PC, PG, R, MC; 2 wk | 24 | FP 100 µg (207) | AZE 274 µg (208) | MP29-02, aka MP-AzeFlu (207) | P (209) | All BID | 18.2 | 18.3 | 18.3 | 18.6 | −5.0 | ± SE | ± SD |

*The paper includes a second study (study 2) comparing montelukast and loratadine, with no INCS arm; the data from study 2, thus, were not appropriate for any of the queries.*

**AM, Morning; AZE, azelastine; BID, twice daily; BDP, beclomethasone dipropionate; BL, baseline; DB, double blind; DD, double dummy; DT, daytime; FP, fluticasone propionate; LCET, levocetirizine; LOR, loratadine; MC, multicenter; MON, montelukast; MOM, mometasone furoate nasal spray; OAH, oral antihistamine; P, placebo; PC, placebo-controlled; PG, parallel group; PM, evening; QD, once daily; R, randomized; SAR, seasonal allergic rhinitis; SD, standard deviation; SE, standard error; Sx, symptom(s); TNSS, total nasal symptoms score; Tx, treatment; VAS, visual analog scale; wk, week(s).**
REFERENCES


