Magnitude of effect of asthma treatments on Asthma Quality of Life Questionnaire and Asthma Control Questionnaire scores: Systematic review and network meta-analysis

Eric D. Bateman, MD,a Dirk Esser, PhD,b Costel Chirila, PhD,c Maria Fernandez, PhD, MBA,c Andy Fowler, MD,d Petra Moroni-Zentgraf, MD,e and J. Mark FitzGerald, MDf Cape Town, South Africa, Ingelheim am Rhein, Germany, Research Triangle Park, NC, Bracknell, United Kingdom, and Vancouver, British Columbia, Canada

Background: The Asthma Quality of Life Questionnaire (AQLQ) and the Asthma Control Questionnaire (ACQ) are widely used in asthma research; however, in studies of newer asthma treatments, mean improvements in these measures compared with placebo arms do not exceed the minimal important difference (MID), particularly when a new treatment is added to current treatment.

Objective: We performed a systematic review and network meta-analysis to examine the magnitude of AQLQ and ACQ responses achieved with commonly used asthma drugs and factors influencing these end points in clinical trials.

Methods: A systematic literature search was conducted to identify blinded randomized controlled trials reporting AQLQ or ACQ results. Mixed treatment comparisons, combined with meta-regression, were then performed.

Results: Of the 64 randomized controlled trials (42,527 patients) identified, 54 included the AQLQ and 11 included the ACQ as end points. The presence of a run-in period, the nature of treatment during the run-in period, concurrent treatment during the treatment period, and instrument version significantly influenced the change in AQLQ score from baseline and whether it exceeded the MID. When compared with placebo, only inhaled corticosteroids (ICSs), with or without a long-acting b-agonist, achieved the MID. The ACQ results were comparable with those of the AQLQ: no differences from placebo exceeded the MID, and ICS-based treatments provided the greatest improvements.

Conclusion: The established within-patient MID for the ACQ and AQLQ is not achievable as a group-wise efficacy threshold between treatment arms in clinical studies in which controllers are added to ICS treatment. Thus in addition to reporting mean changes of the instruments, other measurement criteria should be considered, including responder analyses. (J Allergy Clin Immunol 2015;000:000-000.)

Key words: Asthma, asthma control, quality of life, network meta-analysis, questionnaire

The step-wise approach to pharmacotherapy in patients with asthma involves addition of therapy with the goal of achieving both symptom control and prevention of exacerbations. However, evaluating the efficacy of adding 1 or more controllers to inhaled corticosteroids (ICSs) at step 3 and beyond in the treatment guidelines is challenging for several reasons. Generally, the incremental benefit achieved with each addition is likely to diminish as more treatments are added. Furthermore, because treatments differ in their effect on different asthma end points, improvement might be limited to some, but not all, end points. Demonstrating improvement in patient-reported outcomes (PROs), such as symptom control and health status, is particularly challenging because these outcomes are subjective and can vary widely among subjects. Consequently, their measurement involves the use of complex tools for semiquantitative assessment of symptoms and impairment. Health authorities might require evidence of improvement in 1 or more patient-reported end points for registration, reimbursement, or both of new asthma treatments. Widely used PROs in clinical trials are the Asthma Control Questionnaire (ACQ) and the Asthma Quality of Life Questionnaire (AQLQ). For both instruments, the accepted minimal important difference (MID) is 0.5 units, which is the “smallest difference in score … which patients perceive as beneficial and which would mandate a change in the patient’s management.” The MID is a measure of within-person change but in clinical trials is used as a threshold value for between-group comparisons.
Assessment of recent placebo-controlled studies of new controller treatments in patients with severe asthma (eg, omalizumab and tiotropium) reveals that improvements in AQLQ and ACQ responses are smaller than might be expected. For example, in trials of tiotropium in patients whose symptoms are uncontrolled with at least an ICS and a long-acting \( \beta \)-agonist (LABA), group mean differences in AQLQ and ACQ scores compared with placebo did not exceed the MID for either instrument. These observations call into question the performance of these instruments and the interpretation of results obtained with them, particularly when multiple treatments are being used. It is worth noting that both the AQLQ and the ACQ were developed to assess the impact of new treatments in patients with uncontrolled, symptomatic, or persistent asthma at baseline were included if the overall score changes from baseline values for the AQLQ, ACQ, or both were reported after patients received 1 or more of the following treatments: an ICS, a LABA, a leukotriene receptor antagonist (LTRA), a short-acting \( \beta \)-agonist, omalizumab, or theophylline. Data from all the instrument versions of the AQLQ, such as the Standardized AQLQ and the MiniAQLQ, and of the ACQ, such as the 5-item and 6-item versions (ACQ-5 and ACQ-6), were collected.

Data extraction and assessment of risk of bias

Data from the original studies were extracted by using a standardized abstraction form developed in Microsoft Excel (see the Methods section in this article’s Online Repository for details), which included study design information. To consistently capture key study differences, run-in and background treatments were defined as stable comedations if they were taken by at least 50% of patients in addition to the study medication before randomization, after randomization, or both. Data were independently checked for accuracy by 2 reviewers (C.C. and M.F.): the risk of bias of individual studies was assessed at the study and outcome level by using the quality criteria presented in the National Institute for Health and Care Excellence single technology appraisal guidance and the Centre for Reviews and Dissemination’s guidance (see the Methods section in this article’s Online Repository for details).

Outcome measures

The meta-analysis assessed AQLQ and ACQ score changes from baseline, where baseline was defined as the last visit before the start of the treatment phase. The extracted assessments were based on the time point of the study primary end point, as designated in the publication, or the latest available time point (if no time point was designated as primary).

Statistical methods

For each outcome, a mixed treatment comparison (MTC) combined with meta-regression was performed. Linear mixed models with the SE of mean change from baseline in the instrument used as a weighting variable and trials as random effects were constructed by using the PROC MIXED procedure in SAS (version 9.3; SAS Institute, Cary, NC). If the SE was not available, it was either derived or imputed (see the Methods section in this article’s Online Repository for details). Adjusted least-squares means for each treatment and adjusted mean differences between any 2 treatments, along with 95% CIs, were estimated. Multiple covariates were assessed, both individually and in combination, for inclusion in the MTC model to address heterogeneity and reduce inconsistency between treatment comparisons (see the Methods section in this article’s Online Repository for details). Covariates with \( P \) values of .05 or less were included in the model.

By comparing the estimated mean changes from baseline and their CIs with the MID, the size of the treatment responses were further classified as follows:

- **no effect** if the point estimate did not reach the MID and the 95% CI included zero;
- **no clinically significant effect** if the point estimate did not reach the MID and the 95% CI was between zero and the MID;
- **not significantly less than the MID** if the point estimate did not reach the MID and the 95% CI was greater than zero but contained the MID;
- **probable clinically significant effect** if the point estimate exceeded the MID but the 95% CI contained the MID; and
- **large clinically significant effect** if the point estimate exceeded the MID and the 95% CI exceeded the MID.

For each outcome, a mixed treatment comparison (MTC) combined with meta-regression was performed. Linear mixed models with the SE of mean change from baseline in the instrument used as a weighting variable and trials as random effects were constructed by using the PROC MIXED procedure in SAS (version 9.3; SAS Institute, Cary, NC). If the SE was not available, it was either derived or imputed (see the Methods section in this article’s Online Repository for details). Adjusted least-squares means for each treatment and adjusted mean differences between any 2 treatments, along with 95% CIs, were estimated. Multiple covariates were assessed, both individually and in combination, for inclusion in the MTC model to address heterogeneity and reduce inconsistency between treatment comparisons (see the Methods section in this article’s Online Repository for details). Covariates with \( P \) values of .05 or less were included in the model.

By comparing the estimated mean changes from baseline and their CIs with the MID, the size of the treatment responses were further classified as follows:

- **no effect** if the point estimate did not reach the MID and the 95% CI included zero;
- **no clinically significant effect** if the point estimate did not reach the MID and the 95% CI was between zero and the MID;
- **not significantly less than the MID** if the point estimate did not reach the MID and the 95% CI was greater than zero but contained the MID;
- **probable clinically significant effect** if the point estimate exceeded the MID but the 95% CI contained the MID; and
- **large clinically significant effect** if the point estimate exceeded the MID and the 95% CI exceeded the MID.

**METHODS**

**Search strategy**

A systematic literature search using PubMed, Embase, and the National Health Service Economic Evaluation Database was conducted on April 5, 2012, and updated on June 14, 2013 (details are provided in the Methods section and Table E1 in this article’s Online Repository at www.jacionline.org). In addition, the bibliographies of existing literature reviews and meta-analyses, the clinicaltrials.gov study register, and the 2011 and 2012 conference Web sites of the American Thoracic Society and European Respiratory Society were searched. No limits regarding publication date or language were applied.

**Abbreviations used**

- ACQ: Asthma Control Questionnaire
- AQLQ: Asthma Quality of Life Questionnaire
- ICS: Inhaled corticosteroid
- LABA: Long-acting \( \beta \)-agonist
- LTRA: Leukotriene receptor antagonist
- MID: Minimal important difference
- MTC: Mixed treatment comparison
- PRO: Patient-reported outcome
- RCT: Randomized controlled trial

**Inclusion criteria and selection of studies**

Using predefined inclusion and exclusion criteria, 2 reviewers (C.C. and M.F.) independently scanned titles and abstracts of the identified studies at level 1 screening to evaluate potential study relevance; full texts of studies selected at level 1 were reviewed at level 2 screening (see the Methods section in this article’s Online Repository for full details). Discrepancies were reconciled between the 2 reviewers or by a third reviewer (D.E.), if necessary. Double-blind randomized controlled trials (RCTs) of adolescent and adult patients with uncontrolled, symptomatic, or persistent asthma at baseline were included if the overall score changes from baseline values for the AQLQ, ACQ, or both were reported after patients received 1 or more of the following treatments: an ICS, a LABA, a leukotriene receptor antagonist (LTRA), a short-acting \( \beta \)-agonist, omalizumab, or theophylline. Data from all the instrument versions of the AQLQ, such as the Standardized AQLQ and the MiniAQLQ, and of the ACQ, such as the 5-item and 6-item versions (ACQ-5 and ACQ-6), were collected.
Sensitivity analyses were performed to assess the effect of data imputation for the SE and the effect of including in the analysis studies that compared treatments within the same treatment classes. Inconsistency of the network was also evaluated.19-21 A test result was declared significant if the P value was .05 or less.

RESULTS

Systematic review and summary of included trials

Sixty articles reporting on 64 RCTs (42,527 patients) were identified (see Fig 1 and the Results section in this article’s Online Repository at www.jacionline.org for details).

Table 1 presents a weighted average of the baseline characteristics by treatment. In general, baseline characteristics were similar across treatments.

Overall, the quality of the studies was good (see Table E2 in this article’s Online Repository at www.jacionline.org). Most studies reported the differences in dropouts between the treatment groups and the reasons for patient withdrawal, although reporting of randomization methods, concealment allocation, and blinding of participants was not clear in many of the studies. The risk of bias in the analyses was not clear in several studies because the conduct of intention-to-treat analyses was poorly described and...
applied. For example, the number of patients used in the calculations of health-related quality of life often was not clearly reported.

MTC of AQLQ results

The AQLQ was included as an end point in 54 of the 64 blinded RCTs. Fig 2 shows the AQLQ network diagram; numbered lines indicate the number of studies included in the comparison (see Table E3 in this article’s Online Repository at www.jacionline.org for more detailed information on the individual studies). Higher AQLQ scores indicate better health-related quality of life, such that increases from baseline are considered favorable.

During the run-in period, subjects in 9 (17%) RCTs had no medication (or no run-in period), in 9 (17%) RCTs they received placebo, and in 36 (66%) RCTs they received an ICS. During the treatment phase, subjects in 38 (72%) RCTs had no background treatment, and in 15 (28%) RCTs they continued ICS use as background treatment. Overall, 41 (76%) RCTs used the original AQLQ instrument, whereas 13 (24%) RCTs used alternate versions (Standardized AQLQ [20%] or MiniAQLQ [4%]).

Study treatment, run-in treatment (none [or no run-in period], placebo, or ICS), background treatment (none or ICS), AQLQ type (original or other), and study treatment by background treatment were statistically significant covariates included in the final MTC multivariable model for the AQLQ.

Changes in AQLQ scores from baseline (within-treatment comparisons)

Adjusted AQLQ estimates of change from baseline for pooled treatment groups in trials with or without background medication or with different run-in treatments and type of AQLQ are presented in detail in Table E4 in this article’s Online Repository at www.jacionline.org. Fig 3 shows AQLQ score changes from baseline for situations considered representative of asthma clinical studies by using the AQLQ original version and various combinations of run-in period and background treatment (ie, no run-in treatment and no ICS background, placebo run-in period and no ICS background, ICS run-in period and ICS background).

In trials with no run-in period and no concurrent (background) controller treatment during the treatment period, the estimated AQLQ changes from baseline achieved in the ICS plus LABA, ICS, LTRA, and LABA treatment groups were all greater than 0.5 and categorized as large and clinically significant. However, placebo also achieved a mean improvement greater than 0.5 units (0.553; 95% CI, 0.367-0.739), which was categorized as probably clinically significant.

When a run-in period was present during which placebo was used and there was no concurrent therapy during the active phase of the trials, the estimated AQLQ change from baseline achieved in the ICS plus LABA, ICS, LTRA, and LABA treatment groups were all greater than 0.5 and categorized as large and clinically significant. However, placebo also achieved a mean improvement greater than 0.5 units (0.553; 95% CI, 0.367-0.739), which was categorized as probably clinically significant.

When a run-in period was present during which placebo was used and there was no concurrent therapy during the active phase of the trials, the estimated AQLQ change from baseline was large and clinically significant only for the ICS plus LABA (0.919; 95% CI, 0.734-1.104) and ICS (0.729; 95% CI, 0.549-0.909) treatment groups. For LTRA (0.504; 95% CI, 0.324-0.685), the increase exceeded 0.5 and was of probable clinical significance, but LABA treatment alone (0.419; 95% CI, 0.223-0.616) did not exceed the MID threshold and thus was not clinically significant.

When an ICS formed part of the treatment during the run-in period and was used as background treatment, the estimated AQLQ improvements from baseline for the LABA, LTRA, and omalizumab groups were large and clinically significant; however, those for theophylline, short-acting β-agonist, and placebo were also greater than 0.5 units but only fulfilled the criteria for probable clinical significance (Fig 3).
Comparisons between treatments

Comparison of active versus placebo treatments in patients with and without background ICSs using the same treatment during the run-in period are presented in Fig 4. When background treatment did not include a controller ICS, only ICS plus LABA and ICS alone achieved a mean improvement of greater than 0.5 units. For ICS plus LABA treatment, the increase was large and clinically significant (0.729; 95% CI, 0.658-0.799), but for ICS treatment, the increase was of only probable clinical significance (0.539; 95% CI, 0.479-0.599). LABAs and LTRAs added to concurrent ICS treatment achieved increases of only 0.229 (95% CI, 0.135-0.323) and 0.314 (95% CI, 0.230-0.398), respectively, which were not clinically significant (Fig 4).

When the background concurrent treatment during the active treatment phase of trials was an ICS, improvements in AQLQ scores achieved with active treatments relative to placebo were considerably smaller. With the addition of a LABA (0.349; 95% CI, 0.271-0.427), LTRA (0.198; 95% CI, 0.127-0.269), or omalizumab (0.305; 95% CI, 0.202-0.408), the change was not clinically significant, and theophylline (0.011; 95% CI, −0.198 to 0.219) had no effect (Fig 4).

Adjusted AQLQ relative effect estimates are presented in Table E5 in this article’s Online Repository at www.jacionline.org.

Results of sensitivity analyses were generally similar to the main analysis findings, although the statistical significance of some covariates included in the models changed. The consistency tests showed a disagreement between direct and indirect evidence for the ICS-placebo-LTRA and ICS-LABA-LTRA loops, indicating that covariate adjustment was justified. However, the exact level of significance at which to reject the null hypothesis was no longer .05 because the tested loops and therefore the consistency test were not independent.19

MTC of ACQ results

The ACQ was an end point in 11 of the 64 blinded RCTs. Fig 5 shows the ACQ network diagram; numbered lines indicate the number of studies included in the comparison (see Table E6 in this article’s Online Repository at www.jacionline.org for more detailed information on the individual studies). Lower ACQ scores indicate better control of asthma symptoms, such that decreases from baseline are considered favorable.
During the run-in period, all RCTs included ICS treatment; during the treatment phase, 7 (64%) RCTs had no background treatment, and 4 (36%) RCTs had ICS background treatment. Overall, 7 (64%) RCTs used the ACQ original version (ie, ACQ-7), and 4 (36%) RCTs used the ACQ, 5-item version. With the exception of study treatment, no statistically significant covariates were found for inclusion in the model.

Changes in ACQ scores from baseline (within-treatment comparisons)

Fig 6 presents the unadjusted least-squares means by treatment for the ACQ overall score change from baseline. No studies showed clinically significant changes in ACQ scores. The effects on ACQ least-squares mean changes from baseline were “probably clinically significant” (see the Results section in this article’s Online Repository for details)\(^{18}\) only for the ICS plus LABA pooled treatment group (−0.625; 95% CI, −0.767 to −0.482).

Comparisons between treatments

Fig 7 and Table E7 in this article’s Online Repository at www.jacionline.org present the unadjusted relative treatment effects of ACQ overall score change from baseline versus placebo. None of the pooled treatment groups had effects on ACQ score mean change that exceeded the MID in comparison with placebo change from baseline. Mean reductions in ACQ scores ranged from −0.102 (95% CI, −0.306 to 0.101) for the addition of theophylline to −0.364 (95% CI, −0.471 to −0.257) for the comparison of LABA plus ICS with ICS alone (Fig 7). Thus all estimates revealed either no effect or no clinically significant effect. Results of the sensitivity analyses were similar to those of the main analysis. The results of these analyses should be interpreted with caution because of the small sample size and because no covariate adjustments could be made.

DISCUSSION

Our study provides several important insights into the performance of the AQLQ and ACQ in clinical trials in patients with asthma. First, it confirms that most established asthma therapies achieve a clinically significant improvement in the mean AQLQ score change when compared with baseline values. Second, it confirms that for all active asthma treatments, the placebo effect is greatest when there is no run-in period and when no controller is permitted as concurrent medication during the study period. Third, improvements in AQLQ score in placebo arms are also high in studies when ICSs are used during the run-in period. Thus not surprisingly, compared with placebo treatment arms, only groups receiving ICS plus LABA achieve a clinically significant change, and in our analysis the improvement with ICS achieved only a probably clinically significant level. Finally, when a second controller is added to ICS treatment (ie, in comparisons of placebo vs controller added to patients receiving...
ICSs during the active treatment phase), none of the comparators achieve a clinically significant or even a probably clinically significant result.

The results of our analysis of the ACQ are in line with those for the AQLQ, with none of the differences from placebo exceeding the MID and ICS-based treatments showing the greatest effect. However, this conclusion should be interpreted with caution because of the smaller number of studies suitable for analysis.

This study has several limitations. Reports of trials do not present score values in a consistent manner, and baseline score values were available for only three quarters of the trials. SEs had
to be derived for more than one half of the trials. Several variables (eg, FEV₁ and baseline AQLQ score) could not be used as covariates in the MTC model because of missing values.

Surprisingly, few studies with the ACQ instrument were found in the literature. In fact, only 1 double-blind RCT of omalizumab using the ACQ instrument met our inclusion criteria.22 This might be explained by a reporting bias (ie, underreporting of negative trial results).

A strength of our study was our attempt to include all reports containing usable data. Moreover, the inclusion of data for most accepted controller treatments for asthma provided the opportunity to compare the results of PROs with the accepted general efficacy of different classes of treatment. In general, the relative effects of treatments on AQLQ and ACQ responses were consistent with their place in treatment guidelines.

The results raise several questions regarding use of these widely accepted instruments in clinical studies of asthma. Both have been developed and validated by using standard methodologies and fulfill the characteristics of an evaluative instrument.2,5,6,10 The size (0.5 units) of the MID might be questioned. However, the finding that a 0.5-point change on a 7-point scale is the smallest change that patients can reliably report has been observed with other PROs, such as the Chronic Respiratory Disease Questionnaire.5,18,23 Thus it is rather a limitation and threshold of subjective perception of change than a weakness of the test and appears to be appropriate for within-subject changes. What is in question is the use of the MID for comparisons between patients or groups of patients. Although group mean improvements greater than 0.5 might be highly desirable, our analysis confirms that this is achieved only with a favorable trial design that minimizes the placebo effect and compares ICS or ICS plus LABA with placebo. The prospect of demonstrating a mean incremental benefit of greater than 0.5 through addition of a second or third controller in addition to an ICS, other controllers, or both appears remote. In these trials the Hawthorne effect24 might be particularly pronounced because improved adherence with background treatment can result in improvements in all outcomes from baseline and a reduced opportunity to demonstrate benefit of the new treatment.

Faced with this limitation, how is the benefit of new treatments, particularly treatments that are added to a first- or second-line controller, to be quantified? Guyatt et al7 advised that even small mean differences between treatment and control might have an important effect on patients and that the method for establishing this proportion is a continuous variable suitable for analysis. The value of the change to the patients who benefit and to society or the payer is a different consideration.

The results of our study lend support to the view that these PROs should be presented in a different way, first as a comparison of responder rates or as a net treatment benefit analysis (number of responders less those who deteriorated on the treatment) from which numbers needed to treat can be calculated. Treatment comparisons can be further expressed as a “minimum worthwhile incremental advantage,” which was defined by Jones et al23 as the percentage of patients in a treated population who experience improvement at or greater than the MID, which is considered a worthwhile benefit for the intervention.

Several factors need to be considered in arriving at a minimum worthwhile incremental advantage. A first consideration is patient need. The minimum worthwhile incremental advantage might be small in patients with severe disease (eg, asthma with frequent exacerbations, some of which might be life-threatening) in whom any benefit can be considered worthwhile. Consensus among doctors and discussion with patients can be used to test the value proposition. Considerations of adverse effects and cost to patients and society might also be relevant. For subjective outcomes, discrete-choice modeling techniques might assist in this decision.25,26

Our study highlights the importance of clinical study design if the difference between treatments is under study. It is essential to use methods that reduce the placebo effect seen in patients receiving an ICS as background treatment during clinical trials. These include ensuring clinical stability of asthma control during baseline, longer run-in periods (≥4 weeks), avoidance of changes in medications for extended periods before and during the run-in period, and rigorous assessment of treatment adherence before and after randomization. The use of an electronic dose-monitoring device might be of value in assessing
patient behavior and adherence during run-in and treatment phases. This is of increasing importance in assessing the utility of more expensive treatments in development for severe asthma.

Initiatives such as the Critical Path Institute, which aims to develop new PRO tools for use in asthma research, may lead to instruments that are more responsive to complex interventions, but the threshold for the subjective perception of change might remain a limiting factor.

In conclusion, this network analysis of the magnitude of AQLQ and ACQ responses achieved with commonly used controllers for the treatment of persistent asthma confirms that for most treatments, particularly when controllers are combined, the mean difference between treatment groups exceeding the MID is probably not achievable. Future research should focus on responder analyses and consider an individualized approach based on the value of the invention for those who respond and secondarily for society.

We thank Kate Lothman for careful and helpful edits.

Clinical implications: When compared with placebo, only an ICS, with or without a LABA, achieved the MID. A responder analysis might be more appropriate for expressing the clinical benefit of add-on treatments in patients with asthma.

REFERENCES