

Comparison of scores associating symptoms and rescue medication use for evaluating the efficacy of allergy immunotherapy in seasonal allergic rhinoconjunctivitis: results from five trials

J.-M. Grouin¹, E. Vicaut² and P. Devillier³

¹INSERM 1219, University of Rouen, Rouen, France, ²Clinical Research Unit Saint-Louis Lariboisière Fernand-Widal, University of Paris-Diderot, Paris, France and ³UPRES EA 220, Clinical Research Unit, Airway Diseases Department, Foch Hospital, University of Versailles Saint-Quentin, University Paris Saclay, Suresnes, France

Clinical & Experimental Allergy

Summary

Background Over the past decade, regulatory bodies and scientific societies recommended, as primary efficacy outcome, a score that reflects both symptom severity and use of rescue medication for clinical trials in allergy immunotherapy (AIT).

Objective We sought to compare the results obtained with two subject-specific scores, the Combined Score (CS) and the Adjusted Symptom Score (AdSS), for assessment of AIT in seasonal allergic rhinoconjunctivitis due to birch and grass pollen allergens.

Methods CS and AdSS were evaluated in subjects receiving a 300IR dose of allergen extract daily, by sublingual route, in four clinical trials with the 5-grass pollen tablet (NCT00367640, NCT00409409, NCT00955825 and NCT00418379) and one with the birch pollen solution (NCT01731249). The CS is derived from the Rhinoconjunctivitis Total Symptom Score (RTSS) and the Rescue Medication Score (RMS) giving equal weight to symptoms and medication use. The AdSS is a symptom score adjusting for rescue medication use. Efficacy end-points were analysed using an analysis of covariance linear model.

Results In all trials, despite the different constructs of the two scores, Combined Score or Adjusted Symptom Score were similarly reduced in the 300IR group compared to the placebo group. Treatment effect was consistently demonstrated with both scores, CS and AdSS, used as either daily scores or average of the daily scores over the pollen season. Minor differences with the same statistical conclusions were observed between the results, leading to the same interpretation.

Conclusions and Clinical Relevance The two scores, combined and adjusted scores, for evaluation of clinical efficacy of AIT have led to similar results, with similar statistical conclusions and similar interpretation.

Keywords allergic rhinoconjunctivitis, clinical immunology, immunotherapy and tolerance induction, rhinitis, sublingual immunotherapy, symptom score

Submitted 25 February 2016; revised 22 September 2016; accepted 2 October 2016

Correspondence:

Philippe Devillier, UPRES EA 220, Clinical Research Unit, Airway Diseases Department, Hôpital Foch, University of Versailles Saint-Quentin, University Paris Saclay, 11 rue Guillaume Lenoir, Suresnes 92150, France.

E-mail: p.devillier@hopital-foch.org

Cite this as: J.-M. Grouin, E. Vicaut and P. Devillier, *Clinical & Experimental Allergy*, 2016 (00) 1–10.

Introduction

In allergy immunotherapy (AIT) clinical trials, evidence of treatment efficacy is primarily shown through the improvement in allergy symptoms. For ethical reasons, because of duration of the pollen seasons, trial subjects are permitted to use concomitant rescue medications. As both symptom severity and use of rescue medications are interdependent and can be impacted by AIT, European regulatory authorities and scientific societies

such as the World Allergy Organization (WAO) have advocated that both be reflected in the primary efficacy outcome in clinical trials [1, 2]. Specifically, the WAO task force recommended that symptom scores always be combined with rescue medication scores in a composite score to provide more clinically relevant information on AIT than assessing each of these two outcome parameters separately. This is achievable using a combined score (weighted sum scores of symptoms and rescue medications), or a symptom score adjusted for

the use of medication, all subject-specific, using a pre-defined algorithm [3–6]. However, there is no standardized method for combining symptom and rescue medication scores [7]. Different algorithms have been developed for calculating combined or adjusted scores, but none is universally accepted [2, 8].

In this article, using data from five multi-centre, randomized, double-blind, placebo-controlled, parallel-group trials of sublingual immunotherapy in subjects with seasonal allergic rhinoconjunctivitis (four grass pollen studies and one birch study), we review the results from two scoring methods used to assess efficacy combining symptom scores and rescue medication use: the Combined Score (CS) and the Adjusted Symptom Score (AdSS). These scores were accepted by the US Food and Drug Administration (FDA) and the European Medicines Agency (EMA) as primary or secondary end-points for the clinical trials.

Methods

Trial designs

All trials were conducted in accordance with the principles of the Declaration of Helsinki.

Clinical trials with 5-grass pollen sublingual tablet. Four multi-centre, randomized, double-blind, placebo-controlled, parallel-group trials were conducted between 2004 and 2011 in Europe, Russia and North America investigating the efficacy and safety of the 5-grass pollen tablet. In three studies (European dose-ranging study in adults aged 18 to 45 – NCT00367640; Pediatric study in children and adolescents aged 5 to 17 – NCT00409409; and US study in adults aged 18 to 65 – NCT00955825), subjects with grass pollen-induced allergic rhinoconjunctivitis (ARC) were treated daily with the 5-grass pollen extract sublingual immunotherapy tablet or matching placebo starting about 4 months before the expected start of the grass pollen season and continuing for its duration. In the Long Term study (NCT00418379), adults aged 18–50 received the pre- and coseasonal treatment during three consecutive years (i.e. a discontinuous regimen) and were followed for two subsequent treatment-free periods. Design and efficacy variables that have been previously described in detail [9–13] were similar across studies.

To enable a comprehensive review of the efficacy results, a pooled analysis (meta-analysis on individual subject data) was performed including all subjects who received 300IR (IR for index of reactivity) or placebo beginning 4 months pre-seasonally and continuing throughout the allergy season (i.e. European study, Pediatric study, US study and the first year of the Long Term study).

Clinical trial with birch pollen sublingual solution. This multi-centre, randomized, double-blind, placebo-controlled, parallel-group trial was conducted between 2010 and 2012 in Europe (Birch study – NCT01731249) [14] and evaluated the efficacy and safety of 300IR birch pollen sublingual solution. In this study, eligible subjects with birch pollen-associated ARC aged 18–65 were treated daily with a 300IR sublingual solution of birch pollen allergen extract or placebo initiated approximately 4 months before the expected start of the birch pollen season and continued throughout the season for two pollen seasons.

For all above-mentioned studies, the primary efficacy end-point was in accordance with the regulatory recommendations at the time.

Variables

Independent scoring of symptoms and rescue medications. In the clinical trials of the sublingual tablet of 5-grass pollen extract for AIT of grass pollen-induced ARC, symptom scores and medication use over the previous 24 hours were recorded daily by the subject. Subjects self-assessed six symptoms of ARC (sneezing, rhinorrhea, nasal pruritus, nasal congestion, ocular pruritus and watery eyes) and graded each on a four-point scale: 0 (absence), 1 (mild), 2 (moderate) and 3 (severe). For each subject, the six symptom scores were summed as the Rhinoconjunctivitis Total Symptom Score (RTSS), resulting in a scale from 0 to 18.

In these studies, subjects were permitted to take rescue medication according to a stepwise regimen (step 1: oral or ocular antihistamine, step 2: nasal corticosteroid and step 3: oral corticosteroid) to manage severe rhinoconjunctivitis symptoms at the time of the pollen season. The daily Rescue Medication Score (RMS) was defined by Stallergenes based on the hypothesis that a nasal corticosteroid is more effective than an antihistamine and an oral corticosteroid is more effective than a nasal corticosteroid, leading to a derived ordinal scale: 0 = absent, 1 = antihistamine, 2 = nasal corticosteroid, 3 = oral corticosteroid. In case the subject took two or more rescue medications, the higher score was retained for the corresponding day.

Combined scoring of symptoms and rescue medications. The key concept behind the development of a combined scoring is the need to take into account the impact of rescue medication use on symptoms. In this situation, the observed symptom score is likely to be lower than its counterfactual value (i.e. the one that would have been observed without rescue medication). The CS and the AdSS aim to address this bias.

As advocated by WAO and regulatory authorities since 2006 [1, 2], Stallergenes defined its own CS

derived from the RTSS and the RMS, which gives equal weight to symptoms and medication use. It is calculated for each day for each subject as $CS = (RTSS/6 + RMS)/2$ and is conventionally expressed on a scale from 0 to 3. The FDA has recommended this daily CS as the primary end-point for the US study.

The AdSS [3] is a company-developed symptom score that is adjusted for rescue medication use. The subject evaluated his/her symptoms retrospectively over the previous 24 hours. When s/he took rescue medication during the previous 24 hours, this rescue medication may have had an impact on the symptom score assessment on the day of intake. If the rescue medication is a corticosteroid, the impact on symptoms can be longer than 24 hours [15, 16]. Thus, when a subject took rescue medication, his/her symptom score was adjusted for the current day and the day after. Therefore, the RTSS is adjusted as follows: if the subject took rescue medication on a given day, the AdSS equals the RTSS of that day or the AdSS of the day before, whichever is higher. This adjustment applies to the day of rescue medication use and the following day [3]. Like the RTSS, the AdSS ranges from 0 to 18.

For the purposes of this article, to enable comparison between CS and AdSS expressed on different ranges (0–3 and 0–18, respectively), absolute values are presented. In addition, CS values were multiplied by six to express them on a 0–18 scale. RMS values were also multiplied by six to express them on a 0–18 scale, leading to the calculation of the $CS_{[0-18]}$ as $(RTSS + RMS_{[0-18]})/2$.

Statistics

Evaluation period. For each trial, dates of pollen seasons were defined for each study centre, before data unblinding, using the daily pollen count records in the vicinity of study centres. In grass pollen studies, the start date of the study centre-specific grass pollen season was defined as the first of three consecutive days with a pollen count of 30 grains/m³ (or 10 grains/m³ in the US study) of air or above at that site, and the end date as the last of three consecutive days with a pollen count of 30 grains/m³ (or 10 grains/m³ in the US study) of air or above. In the Birch study, the birch pollen period at each site was defined as the period covering 90% of the total annual birch pollen count.

In the Long Term study, the primary evaluation period was the third grass pollen season (last season on treatment). In the Birch study, the primary evaluation period was the second birch pollen season. Specifics are provided in previous publications [9, 10, 12–14].

Analyses sets. In all trials, the primary analysis sets included all subjects in the selected studies who received at least one dose of the investigational product

and had at least one day with primary efficacy measurement available on treatment during the primary evaluation period.

Analysed variables. As mentioned above, in all trials, symptom scores and rescue medication use were recorded daily by the subjects on daily record cards. The efficacy variables (RTSS, RMS, CS and AdSS) were expressed as: (1) the daily scores recorded each day for each subject during the evaluation period or (2) the average of the daily scores (a summary score), calculated for each subject over the evaluation period [17].

Statistical analyses

Analysis model and covariates. In all trials, the efficacy end-points (i.e. the dependent variables: symptom score, rescue medication score or symptom/rescue medication score) were analysed using an analysis of covariance (ANCOVA) linear model including terms for treatment, pooled centre (pooling was performed according to geographical areas and before data unblinding) and covariates, which could impact the efficacy end-point. The set of adjusted covariates was not identical throughout studies. For the purpose of this article, to ensure the consistency of analyses across studies, analyses were performed adjusting for the same set of covariates (pooled centre, age, gender, asthma status, sensitization status; for the Birch study only, presence of Oral Allergy Syndrome; and study for the pooled analysis).

According to the type of score (average or daily), the following models were fitted:

- Standard linear ANCOVA model for average scores, considering the pollen period as a single time-point [17]
- Linear mixed ANCOVA model for repeated measures for daily scores, using the pollen season day as indicator of time

In the pooled analysis, the interaction between 'study' and 'treatment' was tested to assess the possible heterogeneity of the results across studies.

Least squares (LS) means for treatment groups, adjusted difference vs. placebo along with 95% confidence interval (CI) were estimated. Of note, when the CI rules out the zero value, statistical significance is achieved.

Multiplicity issues. For the European study and Long Term study, due to multiple treatment arms, a step down approach was used to maintain the overall type I error rate at 5%, whereas for two-arm studies, no adjustment for multiplicity was necessary.

All statistical analyses were performed using SAS® software (version 9.2), SAS Institute Inc. Cary, NC, USA.

Results

Average Symptom Score and Medication Score

Over one pollen season. The average scores in the individual 5-grass trials over the primary evaluation period of each study are presented in Table 1. The average scores in the pooled database of these trials over the first pollen season are shown in Table 2. The LS mean differences between the 300IR and placebo groups were statistically significant for all RTSS and RMS_[0–18] (Fig. 1a). These differences ranged from –0.94 to –1.45 for the RTSS and from –0.57 to –1.22 for the RMS_[0–18]. In the pooled analysis, the differences were of –1.02 for the RTSS and –0.91 for the RMS_[0–18].

For the Birch study, the average scores are summarized in Table 4. During the second pollen season of this trial, the LS mean differences between the 300IR and placebo groups were of –1.57 for the RTSS and –1.74 for the RMS_[0–18] and were statistically significant (Fig. 3a).

Over subsequent pollen seasons of the Long Term and Birch studies. The average scores for each of the three

treatment years of the Long Term study are presented in Table 3. The LS mean difference between the 300IR and placebo groups for the symptom scores was statistically significant during the second and third pollen periods (Fig. 2a). Over the treatment period, the differences in RTSS vs. placebo increased year after year. The medication score in the active group was significantly reduced vs. placebo over the years of the study.

In the Birch study, the symptom and medications scores in the active group were also significantly reduced compared to the placebo group, during the first and second pollen seasons (Table 4 and Fig. 3a).

Average Combined Score and Adjusted Symptom Score

Over one pollen season. As shown in Figs 1b and 3b, the reductions of the two scores, CS_[0–18] and AdSS, in the active group were significant compared to placebo.

Over pollen seasons of the Long Term and Birch studies. In the grass Long Term and Birch studies, CS_[0–18] and AdSS assessed every year were significantly reduced in the 300IR groups compared to the placebo

Table 1. Analysis of average scores in individual clinical trials with 5-grass pollen tablet (primary analysis sets)*

Treatment	<i>n</i>	RTSS [0–18]	RMS [†] [0–18]	CS ^{†,‡} [0–18]	AdSS [0–18]
European study					
300IR LS mean (SE)	136	3.57 (0.309)	2.36 (0.310)	2.97 (0.253)	4.23 (0.364)
Placebo LS mean (SE)	148	5.02 (0.307)	3.40 (0.308)	4.22 (0.251)	6.06 (0.361)
Pediatric study					
300IR LS mean (SE)	131	3.30 (0.293)	3.17 (0.369)	3.23 (0.279)	4.17 (0.379)
Placebo LS mean (SE)	135	4.43 (0.288)	4.39 (0.363)	4.39 (0.274)	5.81 (0.372)
US study					
300IR LS mean (SE)	208	3.64 (0.314)	0.84 (0.166)	2.24 (0.201)	3.98 (0.346)
Placebo LS mean (SE)	228	4.58 (0.299)	1.41 (0.158)	3.00 (0.191)	5.22 (0.329)
Long Term study – Year 3					
300IR LS mean (SE)	149	2.67 (0.297)	1.85 (0.251)	2.27 (0.242)	3.39 (0.357)
Placebo LS mean (SE)	165	4.03 (0.289)	2.80 (0.244)	3.42 (0.236)	5.21 (0.348)

n, Number of evaluable subjects for the statistical model; RTSS, Rhinoconjunctivitis Total Symptom Score; RMS, Rescue Medication Score; CS, Combined Score; AdSS, Adjusted Symptom Score; IR, index of reactivity; LS, least squares; SE, standard error; CI, confidence interval.

ANCOVA with pooled site, age, gender, asthma status and sensitization status as covariates.

*The primary evaluation period was the single pollen season for European, Pediatric and US studies and the Year 3 pollen season for the Long Term study.

[†]For ease of comparison between scores, CS and RMS results were recalculated on a 0–18 scale in this article.

[‡]CS_[0–18] was calculated as (RTSS + RMS_[0–18])/2 in this article.

Table 2. Analysis of average scores in pooled data from four clinical trials – 4-month pre-seasonal treatment with 5-grass pollen tablet (primary analysis set)

Treatment	<i>n</i>	RTSS [0–18]	RMS* [0–18]	CS* [†] [0–18]	AdSS [0–18]
300IR LS mean (SE)	663	3.74 (0.150)	2.40 (0.135)	3.07 (0.116)	4.52 (0.180)
Placebo LS mean (SE)	716	4.76 (0.147)	3.31 (0.132)	4.03 (0.113)	6.04 (0.176)

n, Number of evaluable subjects for the statistical model; RTSS, Rhinoconjunctivitis Total Symptom Score; RMS, Rescue Medication Score; CS, Combined Score; AdSS, Adjusted Symptom Score; IR, index of reactivity; LS, least squares; SE, standard error; CI, confidence interval.

ANCOVA with pooled site, age, gender, asthma status, sensitization status and study as covariates.

*For ease of comparison between scores, CS and RMS results were recalculated on a 0–18 scale in this article.

[†]CS_[0–18] was calculated as (RTSS + RMS_[0–18])/2 in this article.

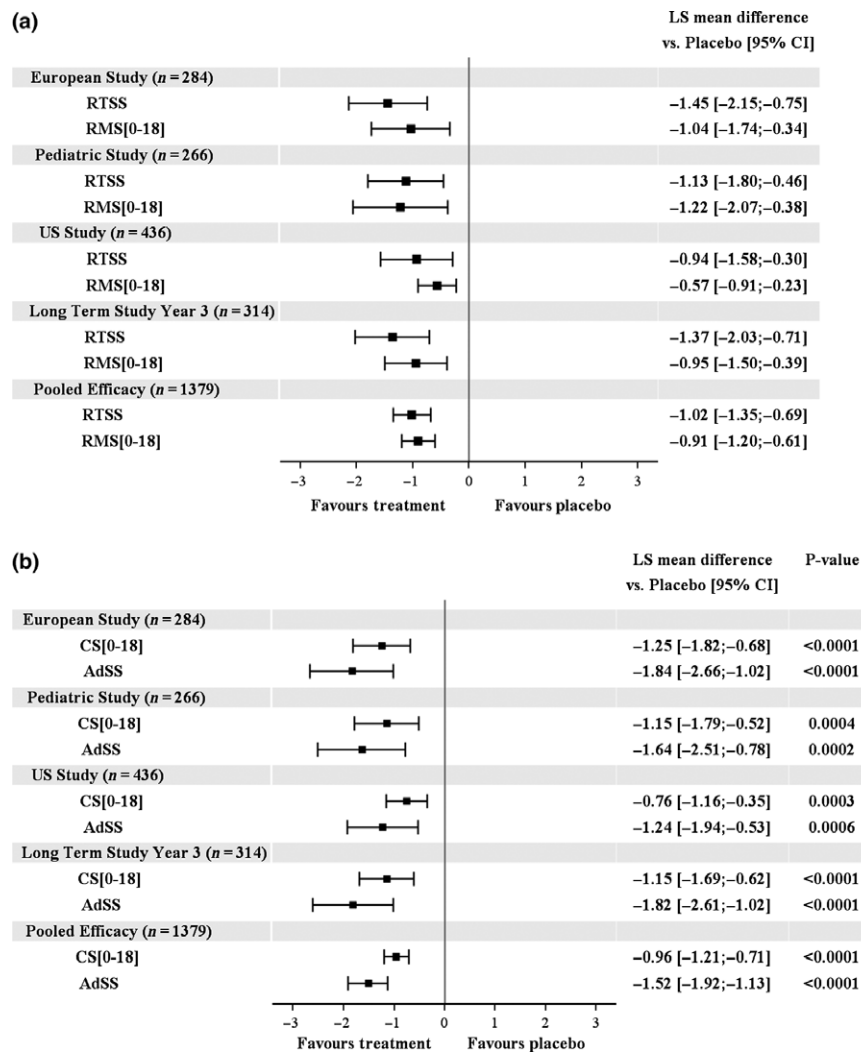


Fig. 1. Forest plot of average scores (a: RTSS and RMS; b: CS and AdSS) in individual and pooled data from four clinical trials – 4-month pre-seasonal treatment with 5-grass pollen tablet (primary analysis sets). n, Number of evaluable subjects; LS, least squares; CI, confidence interval; RTSS, Rhinoconjunctivitis Symptom Score; RMS, Rescue Medication Score; CS, Combined Score; AdSS, Adjusted Symptom Score.

groups and were also consistently reduced over the pollen seasons from year to year (Tables 3 and 4; Figs 2b and 3b).

Overall, in all studies, the CS_[0-18] ranged from 2.24 to 5.53 and the AdSS ranged from 3.39 to 7.56, both scores being assessed on a 0–18 scale (Tables 1–4).

Daily Symptom Score, Medication Score, Combined Score and Adjusted Symptom Score

For ease of reading, the daily score results are provided in Supporting Information (Tables S1–S4). For all efficacy variables, daily and average score results were found to be similar over one pollen season or over pollen seasons of the Long Term and Birch studies. Ample evidence of this is shown in the values of the LS mean difference vs. placebo (95% CI) provided in Figs 1–3 vs. the corresponding values in Tables S1–S4.

Heterogeneity test in the pooled analysis

The heterogeneity test performed in each pooled analysis was not statistically significant, which confirms the consistency between the 5-grass trials (data not shown).

Discussion

Our purpose was to compare the results obtained with the two scoring methods: the CS, a composite score, and the non-composite AdSS, for evaluation of the efficacy of sublingual immunotherapy in seasonal aeroallergen-induced ARC in line with the recommendations from regulatory authorities and clinical experts [1, 2]. Since 2004, as the guidelines for the development of new products for the treatment of allergic rhinitis have evolved, so have the primary efficacy end-points of the clinical trials. In 2004, the guidelines recommended as

Table 3. Analysis of average scores in Long Term study – 4-month pre-seasonal treatment with 5-grass pollen tablet (primary analysis sets)*

Treatment	<i>n</i>	RTSS [0–18]	RMS [†] [0–18]	CS ^{†,‡} [0–18]	AdSS [0–18]
Long Term study – Year 1					
300IR LS mean (SE)	188	4.49 (0.244)	3.61 (0.259)	4.04 (0.208)	5.74 (0.318)
Placebo LS mean (SE)	205	4.97 (0.234)	4.48 (0.249)	4.71 (0.199)	6.99 (0.305)
Long Term study – Year 2					
300IR LS mean (SE)	160	3.22 (0.278)	2.11 (0.255)	2.67 (0.230)	4.03 (0.341)
Placebo LS mean (SE)	172	4.42 (0.275)	3.50 (0.252)	3.96 (0.228)	5.92 (0.338)
Long Term study – Year 3					
300IR LS mean (SE)	149	2.67 (0.297)	1.85 (0.251)	2.27 (0.242)	3.39 (0.357)
Placebo LS mean (SE)	165	4.03 (0.289)	2.80 (0.244)	3.42 (0.236)	5.21 (0.348)

n, Number of evaluable subjects for the statistical model; RTSS, Rhinoconjunctivitis Total Symptom Score; RMS, Rescue Medication Score; CS, Combined Score; AdSS, Adjusted Symptom Score; IR, index of reactivity; LS, least squares; SE, standard error; CI, confidence interval.

ANCOVA with pooled site, age, gender, asthma status and sensitization status as covariates.

*The primary evaluation period for this study was the Year 3 pollen season.

[†]For ease of comparison between scores, CS and RMS results were recalculated on a 0–18 scale in this article.

[‡]CS_[0–18] was calculated as (RTSS + RMS_[0–18])/2 in this article.

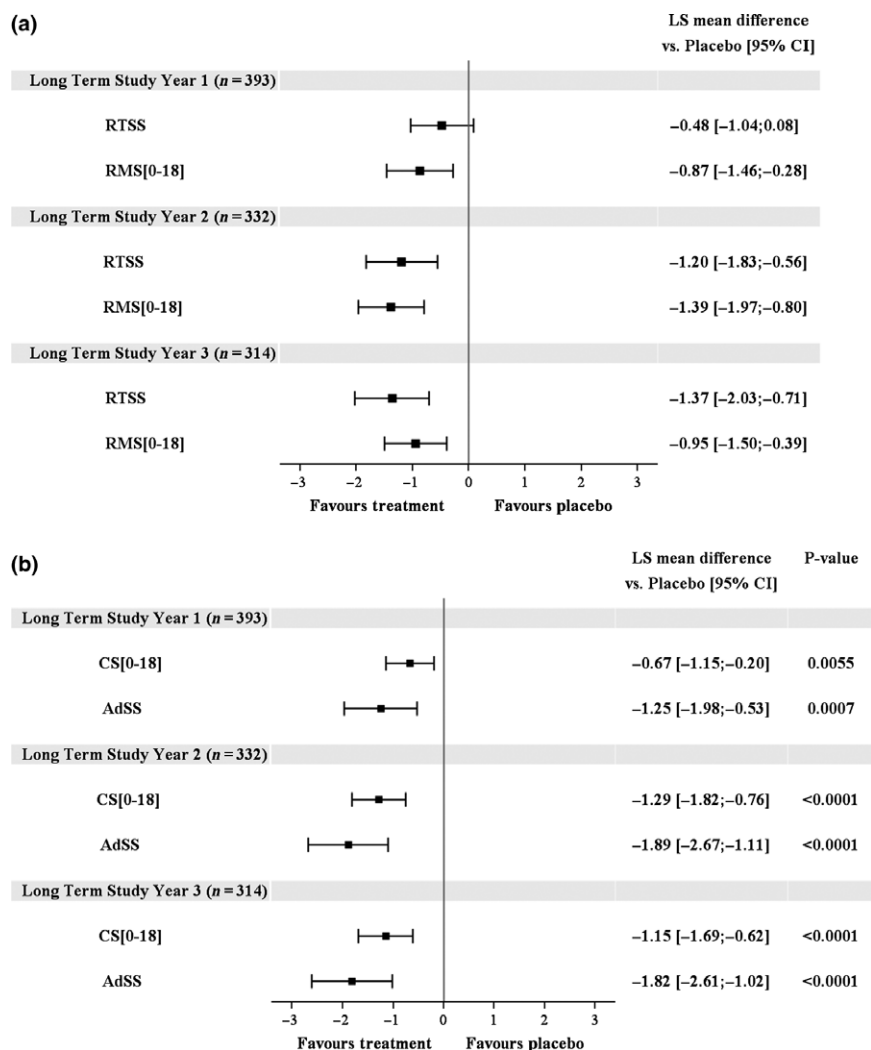


Fig. 2. Forest plot of average scores (a: RTSS and RMS; b: CS and AdSS) in Long Term study – 4-month pre-seasonal treatment with 5-grass pollen tablet (primary analysis sets). *n*, Number of evaluable subjects; LS, least squares; CI, confidence interval; RTSS, Rhinoconjunctivitis Symptom Score; RMS, Rescue Medication Score; CS, Combined Score; AdSS, Adjusted Symptom Score.

Table 4. Analysis of average scores in a clinical trial with birch pollen extract solution (primary analysis sets)*

Treatment	n	RTSS [0–18]	RMS [†] [0–18]	CS ^{†,‡} [0–18]	AdSS [0–18]
Birch study – Year 1					
300IR LS mean (SE)	261	5.19 (0.233)	3.48 (0.260)	4.33 (0.215)	6.09 (0.274)
Placebo LS mean (SE)	275	6.07 (0.244)	4.98 (0.272)	5.53 (0.225)	7.56 (0.287)
Birch study – Year 2					
300IR LS mean (SE)	247	4.04 (0.238)	2.37 (0.237)	3.20 (0.206)	4.64 (0.274)
Placebo LS mean (SE)	253	5.62 (0.250)	4.10 (0.250)	4.86 (0.217)	6.70 (0.289)

n, Number of evaluable subjects for the statistical model; RTSS, Rhinoconjunctivitis Total Symptom Score; RMS, Rescue Medication Score; CS, Combined Score; AdSS, Adjusted Symptom Score; IR, index of reactivity; LS, least squares; SE, standard error.

ANCOVA with pooled site, age, gender, asthma status, sensitization status and oral allergy syndrome as covariates.

*The primary evaluation period for this study was the Year 2 pollen season.

[†]For ease of comparison between scores, CS and RMS results were recalculated on a 0–18 scale in this article.

[‡]CS_[0–18] was calculated as (RTSS + RMS_[0–18])/2 in this article.

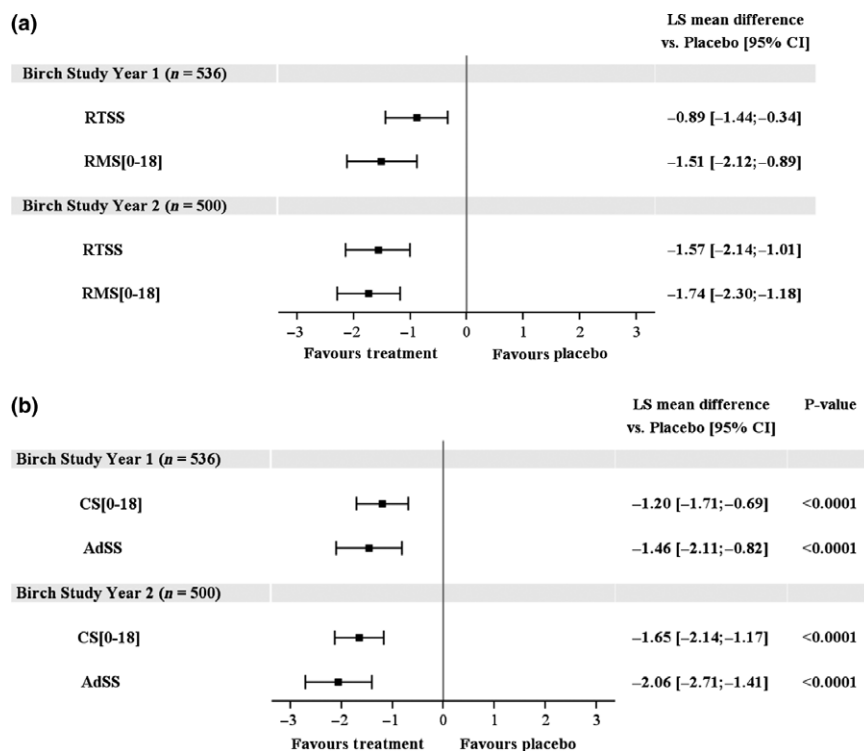


Fig. 3. Forest plot of average scores (a: RTSS and RMS; b: CS and AdSS) in a clinical trial with birch pollen extract solution (primary analysis sets). n, Number of evaluable subjects; LS, least squares; CI, confidence interval; RTSS, Rhinoconjunctivitis Symptom Score; RMS, Rescue Medication Score; CS, Combined Score; AdSS, Adjusted Symptom Score.

the primary variable a total score of rhinitis and conjunctivitis symptoms with each symptom severity assessed on a four-point scale from 0 to 3 [18]. Therefore, the RTSS was the primary variable for the so-called European and Pediatric studies. In 2008, new EMA guidelines for allergy immunotherapy products and the WAO recommended the use of a primary variable reflecting the treatment effect on both symptoms and the use of symptomatic medications [1, 2]. AdSS was developed by Stallergenes and submitted to the EMA with a clinical justification for this end-point, in compliance with the draft guideline

(CHMP/EWP/18504/06). The agency considered this approach 'a suitable primary endpoint taking into account the symptom severity and the influence of the rescue medication' and it was used as the primary end-point of the grass Long Term study. By contrast, for the US study, the FDA recommended the other approach to combining symptom and rescue medication scores, the CS.

Both scores have their merits. Although, they assess different constructs, each aims at compensating for the expected decrease in symptom severity caused by the intake of rescue medication.

The CS is a combination of two scores that assess two separate constructs (i.e. symptom severity and rescue medication grade). The metrics underlying each score are not necessarily consistent (i.e. one point on the symptom severity scale is not necessarily comparable to one point on the rescue medication scale). Hence, the CS relies on a metric that is a mix of two heterogeneous metrics and cannot be considered as a symptom score. Conversely, the AdSS remains a symptom score that assesses the same construct as its unadjusted counterpart, the RTSS.

In all studies, the CS absolute values in the placebo groups were consistently lower than those of AdSS. On one hand, the CS_[0-18] takes into account a symptom score lowered by the rescue medication intake and a modest rescue medication score. On the other hand, the AdSS represents an increase in the RTSS to compensate for the impact of rescue medication intake and therefore reflects the true severity of symptoms.

The review of the absolute differences showed that those estimated with AdSS are greater than those estimated with CS_[0-18]. Concomitantly, the variability of these estimates, as expressed by the confidence intervals (CIs), is greater for AdSS than for CS_[0-18] (Figs 1b, 2b and 3b). In the end, whether CS or AdSS, the score differences between active treatment and placebo can be compared and effect sizes obtained from both CS and AdSS can be considered similar. Furthermore, when the intake of rescue medication is high due to sudden severe symptoms as observed in the Birch study [RMS LS mean in the placebo group = 4.98 and 4.10 at Year 1 and Year 2, respectively (Table 4) vs. 3.31 in the placebo group of the grass studies pooled analysis (Table 2)], the variability of the CS_[0-18] is greater (Fig. 3a,b). Conversely, when the consumption of rescue medication is low as in grass pollen studies and more specifically in the US study, the CIs of the CS_[0-18] are narrower and CS_[0-18] and AdSS results are closer (Figs 1a,b and 2a,b).

The relative differences per se need to be seen in the context with the numeric size of the score on which the relative differences are based [19]. Those are driven by the level of the score in the placebo group, which varies over time and over geographical locations. The apparent greater contrast with AdSS is related to the higher absolute values but is not different if expressed as relative values.

Given the impact of rescue medication use on symptom scores [5, 20], the method for evaluating this use is an important factor to take into consideration. The CS accounts for the daily intake of rescue medication (i.e. for each day of consumption, refer to Methods Section). Alternatively, with the conservative approach assigning a two-day duration of relief whatever the nature of the

rescue medication, the AdSS adjusts the symptom score for 2 days, which may overestimate the weight of the rescue medication [21]. In addition, precise doses of rescue medication are not accounted for in either of the two scores.

Notably, the medication scores are not standardized across companies and different scoring scales are used [6, 20]. Stallergenes has chosen to use an identical rescue medication scoring for all allergens. This RMS is the one recently recommended in the EAACI position paper [5].

Although the CS and AdSS evaluate distinct constructs, both were consistently reduced in actively treated subjects compared to placebo-recipients with no meaningful differences between the scores, either daily or average, as shown by overlapping CIs (Figs 1b, 2b and 3b and Tables S1–S4). This consistency between the two scores was observed on various situations with two different allergens (5-grass pollen or birch), in adult and paediatric population and in Europe or USA, leading to the same conclusions whatever the score.

Conclusions

Treatment effect with grass pollen and birch allergen extracts administered at a 300IR dose with a pre- and coseasonal regimen is consistently demonstrated with the two scores, CS and AdSS. While the AdSS was considered suitable by EMA, the CS was also acknowledged by the WAO and US FDA [1, 22].

The results of all studies whether conducted in Europe or the USA, whether in adults or children, whether in grass or birch allergy and whether analysed as daily or average score over the season are consistent regardless of the score (composite or adjusted). For demonstration of treatment efficacy, both scores have been considered relevant by regulatory authorities and the World Allergy Organization [8]. Recently, the use of the Combined Score has been selected by the clinical experts of the EAACI Task Force as the composite score for demonstration of the efficacy in AIT [5].

Although companies use different rescue medication scoring systems and different combinations of symptom and rescue medication scores, all are justified. Nevertheless, allergen exposure varies depending on the geographical location, the patient's way of life and from year to year. Therefore, the use of those scores, even standardized, will not enable a direct comparison between clinical trials conducted with aeroallergens. Additional work should be performed to further investigate whether both scores still yield similar results in rhinoconjunctivitis induced by allergens other than seasonal allergen.

Acknowledgements

Medical writing and editorial assistance was provided by Josiane Cognet-Sicé and Kathy Abiteboul, biostatistical assistance was provided by Armelle Montagut, all employees of Stallergenes Greer. We thank Dr. Robert K. Zeldin for his contribution in the programme.

Funding sources

Support for this manuscript was provided by Stallergenes Greer, Antony, France.

Conflict of interest

Jean-Marie Grouin has received consulting fees from Stallergenes Greer.

Eric Vicaut has acted as a paid consultant to Stallergenes Greer and has received funding for research carried out in this work and honoraria for advisory board/lectures from AstraZeneca, Bayer, Bristol-Myers Squibb, Daiichi-Sankyo, Eli Lilly, Johnson & Johnson, Merck Sharpe and Dohme, and Pfizer.

Philippe Devillier has received consulting fees, honoraria for lectures and/or research funding from Schering-Plough-MSD, Menarini, GlaxoSmithKline, AstraZeneca, ALK and Stallergenes Greer.

References

- 1 Canonica GW, Baena-Cagnani CE, Bousquet J *et al.* Recommendations for standardization of clinical trials with allergen specific immunotherapy for respiratory allergy. A statement of a World Allergy Organization (WAO) taskforce. *Allergy* 2007; **62**:317–24.
- 2 EMA, European Medicines Agency. CHMP/EWP/18504/2006. *Guideline on the clinical development of products for specific immunotherapy for the treatment of allergic diseases*. London: Committee for Medicinal Products for Human Use (CHMP), November 2008.
- 3 Grouin JM, Vicaut E, Jean-Alphonse S *et al.* The average Adjusted Symptom Score, a new primary efficacy endpoint for specific allergen immunotherapy trials. *Clin Exp Allergy* 2011; **41**:1282–8.
- 4 Hafner D, Reich K, Matricardi PM *et al.* Prospective validation of 'Allergy-Control-SCORE™': a novel symptom-medication score for clinical trials. *Allergy* 2011; **66**:629–36.
- 5 Pfaar O, Demoly P, Gerth van Wijk R *et al.* Recommendations for the standardization of clinical outcomes used in allergen immunotherapy trials for allergic rhinoconjunctivitis: an EAACI Position Paper. *Allergy* 2014; **69**:854–67.
- 6 Pfaar O, Kleine-Tebbe J, Hörmann K *et al.* Allergen-specific immunotherapy: which outcome measures are useful in monitoring clinical trials? *Immunol Allergy Clin North Am* 2011; **31**:289–309.
- 7 Bousquet J, Schünemann HJ, Bousquet PJ *et al.* How to design and evaluate randomized controlled trials in immunotherapy for allergic rhinitis: an ARIA-GA(2) LEN statement. *Allergy* 2011; **66**:765–74.
- 8 Canonica GW, Cox LS, Pawankar R *et al.* Sublingual immunotherapy: World Allergy Organization position paper 2013 update. *World Allergy Organ J* 2014; **7**:6.
- 9 Cox LS, Casale Tb, Nayak AS *et al.* Clinical efficacy of 300IR 5-grass pollen sublingual tablet in a US study: the importance of allergen-specific serum IgE. *J Allergy Clin Immunol* 2012; **130**:1327–34.e1.
- 10 Didier A, Malling HJ, Worm M *et al.* Optimal dose, efficacy, and safety of once-daily sublingual immunotherapy with a 5-grass pollen tablet for seasonal allergic rhinitis. *J Allergy Clin Immunol* 2007; **120**:1338–45.
- 11 Didier A, Malling HJ, Worm M *et al.* Post-treatment efficacy of discontinuous treatment with 300IR 5-grass pollen sublingual tablet in adults with grass pollen-induced allergic rhinoconjunctivitis. *Clin Exp Allergy* 2013; **43**:568–77.
- 12 Didier A, Worm M, Horak F *et al.* Sustained 3-year efficacy of pre- and coseasonal 5-grass-pollen sublingual immunotherapy tablets in patients with grass pollen-induced rhinoconjunctivitis. *J Allergy Clin Immunol* 2011; **128**:559–66.
- 13 Wahn U, Tabar A, Kuna P *et al.* Efficacy and safety of 5-grass-pollen sublingual immunotherapy tablets in pediatric allergic rhinoconjunctivitis. *J Allergy Clin Immunol* 2009; **123**:160–6.
- 14 Worm M, Rak S, De Blay F *et al.* Sustained efficacy and safety of a 300IR daily dose of a sublingual solution of birch pollen allergen extract in adults with allergic rhinoconjunctivitis: results of a double-blind, placebo-controlled study. *Clin Transl Allergy* 2014; **4**:7.
- 15 Derendorf H, Meltzer EO. Molecular and clinical pharmacology of intranasal corticosteroids: clinical and therapeutic implications. *Allergy* 2008; **63**:1292–300.
- 16 Meltzer EO, Blaiss MS, Naclerio RM *et al.* Burden of allergic rhinitis: allergies in America, Latin America, and Asia-Pacific adult surveys. *Allergy Asthma Proc* 2012; **33**:S113–41.
- 17 Senn S, Stevens L, Chaturvedi N. Repeated measures in clinical trials: simple strategies for analysis using summary measures. *Stat Med* 2000; **19**:861–77.
- 18 EMA, European Medicines Agency. CHMP/EWP/2455/02. *Guideline on the clinical development of medicinal products for the treatment of allergic rhinoconjunctivitis*. London: Committee for Medicinal Products for Human Use (CHMP), October 2004.
- 19 Demoly P, Emminger W, Rehm D *et al.* Effective treatment of house dust mite-induced allergic rhinitis with 2 doses of the SQ HDM SLIT-tablet: results from a randomized double-blind, placebo-controlled phase III trial. *J Allergy Clin Immunol* 2015; **137**:444–51.
- 20 Calderon MA, Bernstein DI, Blaiss M *et al.* A comparative analysis of symptom and medication scoring methods used in clinical trials of sublingual immunotherapy for seasonal allergic rhinitis. *Clin Exp Allergy* 2014; **44**:1228–39.

- 21 Didier A, Malling HJ, Worm M *et al*. Prolonged efficacy of the 300IR 5-grass pollen tablet up to 2 years after treatment cessation, as measured by a recommended daily combined score. *Clin Transl Allergy* 2015; 5:12.
- 22 FDA, December 11, 2013: Allergenic Products Advisory Committee Meeting Transcript. *Biologic License Application for sweet vernal, orchard, perennial rye, timothy and Kentucky blue grass mixed pollens allergen extract tablet for sublingual use*. Accessed at: www.fda.gov/AdvisoryCommittee. Meeting Materials Allergenic Products Advisory Committee (APAC), 2013.

Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article:

Table S1. Analysis of daily scores in individual clinical trials with 5-grass pollen tablet (Primary analysis sets)*.

Table S2. Analysis of daily scores in pooled data from 4 clinical trials – 4-month pre-seasonal treatment with 5-grass pollen tablet (Primary analysis set).

Table S3. Analysis of daily scores in Long Term study – 4-month pre-seasonal treatment with 5-grass pollen tablet (Primary analysis sets)*.

Table S4. Analysis of daily scores in a clinical trial with birch pollen extract solution (Primary analysis set)*.

Supplementary Table 1. Analysis of daily scores in individual clinical trials with 5-grass pollen tablet (Primary analysis sets)*

Treatment	n	RTSS [0–18]	RMS [†] [0–18]	CS [‡] [0–18]	AdSS [0–18]
European study					
300IR LS mean (SE)	136	3.48 (0.309)	2.46 (0.311)	2.97 (0.254)	4.30 (0.365)
Placebo LS mean (SE)	148	4.91 (0.307)	3.51 (0.309)	4.22 (0.252)	6.14 (0.362)
LS mean difference vs. placebo [95% CI]		-1.44 [-2.13; -0.74]	-1.05 [-1.75; -0.36]	-1.25 [-1.82; -0.68]	-1.84 [-2.66; -1.02]
Pediatric study					
300IR LS mean (SE)	131	2.52 (0.269)	2.87 (0.382)	2.63 (0.285)	3.65 (0.384)
Placebo LS mean (SE)	135	3.63 (0.265)	4.06 (0.375)	3.77 (0.280)	5.27 (0.378)
LS mean difference vs. placebo [95% CI]		-1.11 [-1.71; -0.51]	-1.19 [-2.03; -0.35]	-1.13 [-1.77; -0.50]	-1.62 [-2.49; -0.76]
US study					
300IR LS mean (SE)	208	3.21 (0.315)	0.63 (0.167)	1.92 (0.201)	3.48 (0.347)
Placebo LS mean (SE)	228	4.16 (0.299)	1.21 (0.159)	2.68 (0.191)	4.73 (0.330)
LS mean difference vs. placebo [95% CI]		-0.95 [-1.59; -0.31]	-0.58 [-0.91; -0.24]	-0.76 [-1.16; -0.35]	-1.25 [-1.95; -0.54]
Long Term study – Year 3					
300IR LS mean (SE)	149	2.18 (0.287)	1.47 (0.229)	1.84 (0.227)	2.83 (0.361)
Placebo LS mean (SE)	165	3.55 (0.280)	2.40 (0.224)	2.99 (0.222)	4.64 (0.352)
LS mean difference vs. placebo [95% CI]		-1.37 [-1.99; -0.74]	-0.93 [-1.41; -0.44]	-1.15 [-1.63; -0.66]	-1.81 [-2.60; -1.02]

n, number of evaluable subjects for the statistical model; RTSS, Rhinoconjunctivitis Total Symptom Score; RMS, Rescue Medication Score; CS, Combined Score; AdSS, Adjusted Symptom Score; IR, Index of reactivity; LS, Least Squares; SE, Standard Error; CI, Confidence interval

ANCOVA with pooled site, age, gender, asthma status and sensitization status as covariates

*The primary evaluation period was the single pollen season for European, Pediatric and US studies and the Year 3 pollen season for the Long Term study.

[†]For ease of comparison between scores, CS and RMS results were recalculated on a 0–18 scale in this article.

[‡]CS_[0–18] was calculated as (RTSS + RMS_[0–18])/2 in this article.

Supplementary Table 2. Analysis of daily scores in pooled data from 4 clinical trials - 4-month pre-seasonal treatment with 5-grass pollen tablet (Primary analysis set)

Treatment	n	RTSS [0–18]	RMS* [0–18]	CS* [†] [0–18]	AdSS [0–18]
300IR LS mean (SE)	663	2.98 (0.158)	1.93 (0.146)	2.44 (0.112)	3.68 (0.188)
Placebo LS mean (SE)	716	4.00 (0.154)	2.83 (0.143)	3.39 (0.110)	5.21 (0.184)
LS mean difference vs. placebo [95% CI]		-1.02 [-1.34; -0.69]	-0.90 [-1.19; -0.61]	-0.95 [-1.17; -0.73]	-1.52 [-1.91; -1.13]

n, number of evaluable subjects for the statistical model; RTSS, Rhinoconjunctivitis Total Symptom Score; RMS, Rescue Medication Score; CS, Combined Score; AdSS, Adjusted Symptom Score; IR, Index of reactivity; LS, Least Squares; SE, Standard Error; CI, Confidence interval

ANCOVA with pooled site, age, gender, asthma status, sensitization status and study as covariates

*For ease of comparison between scores, CS and RMS results were recalculated on a 0–18 scale in this article.

[†]CS_[0–18] was calculated as (RTSS + RMS_[0–18])/2 in this article.

Supplementary Table 3. Analysis of daily scores in Long Term study - 4-month pre-seasonal treatment with 5-grass pollen tablet (Primary analysis sets)*

Treatment	n	RTSS [0–18]	RMS [†] [0–18]	CS ^{†‡} [0–18]	AdSS [0–18]
Long Term study – Year 1					
300IR LS mean (SE)	188	3.84 (0.250)	2.95 (0.267)	3.38 (0.197)	4.87 (0.324)
Placebo LS mean (SE)	205	4.31 (0.240)	3.80 (0.257)	4.04 (0.190)	6.12 (0.311)
LS mean difference vs. placebo [95% CI]		-0.48 [-1.03; 0.08]	-0.85 [-1.44; -0.26]	-0.66 [-1.09; -0.24]	-1.25 [-1.97; -0.53]
Long Term study – Year 2					
300IR LS mean (SE)	160	2.61 (0.000)	1.57 (0.267)	2.09 (0.238)	3.20 (0.332)
Placebo LS mean (SE)	172	3.81 (0.000)	2.94 (0.264)	3.37 (0.235)	5.08 (0.328)
LS mean difference vs. placebo [95% CI]		-1.19 [-1.83; -0.56]	-1.37 [-1.95; -0.79]	-1.28 [-1.80; -0.76]	-1.88 [-2.61; -1.15]
Long Term study – Year 3					
300IR LS mean (SE)	149	2.18 (0.287)	1.47 (0.229)	1.84 (0.227)	2.83 (0.361)
Placebo LS mean (SE)	165	3.55 (0.280)	2.40 (0.224)	2.99 (0.222)	4.64 (0.352)
LS mean difference vs. placebo [95% CI]		-1.37 [-1.99; -0.74]	-0.93 [-1.41; -0.44]	-1.15 [-1.63; -0.66]	-1.81 [-2.60; -1.02]

n, number of evaluable subjects for the statistical model; RTSS, Rhinoconjunctivitis Total Symptom Score; RMS, Rescue Medication Score; CS, Combined Score; AdSS, Adjusted Symptom Score; IR, Index of reactivity; LS, Least Squares; SE, Standard Error; CI, Confidence interval

ANCOVA with pooled site, age, gender, asthma status and sensitization status as covariates

*The primary evaluation period for this study was the Year 3 pollen season.

[†]For ease of comparison between scores, CS and RMS results were recalculated on a 0–18 scale in this article.

[‡]CS_[0–18] was calculated as (RTSS + RMS_[0–18])/2 in this article.

Supplementary Table 4. Analysis of daily scores in a clinical trial with birch pollen extract solution (Primary analysis set)*

Treatment	n	RTSS [0–18]	RMS [†] [0–18]	CS ^{†‡} [0–18]	AdSS [0–18]
Birch study – Year 1					
300IR LS mean (SE)	261	5.26 (0.237)	3.75 (0.266)	4.51 (0.219)	6.42 (0.273)
Placebo LS mean (SE)	275	6.15 (0.248)	5.26 (0.278)	5.70 (0.229)	7.88 (0.285)
LS mean difference vs. placebo [95% CI]		-0.89 [-1.44; -0.33]	-1.51 [-2.13; -0.90]	-1.20 [-1.71; -0.69]	-1.46 [-2.09; -0.82]
Birch study – Year 2					
300IR LS mean (SE)	247	3.96 (0.229)	2.59 (0.252)	3.27 (0.203)	4.68 (0.283)
Placebo LS mean (SE)	253	5.50 (0.239)	4.27 (0.263)	4.89 (0.211)	6.71 (0.296)
LS mean difference vs. placebo [95% CI]		-1.55 [-2.06; -1.03]	-1.68 [-2.24; -1.12]	-1.61 [-2.06; -1.16]	-2.03 [-2.68; -1.39]

n, number of evaluable subjects for the statistical model; RTSS, Rhinoconjunctivitis Total Symptom Score; RMS, Rescue Medication Score; CS, Combined Score; AdSS, Adjusted Symptom Score; IR, Index of reactivity; LS, Least Squares; SE, Standard Error; CI, Confidence interval

ANCOVA with pooled site, age, gender, asthma status, sensitization status and oral allergy syndrome as covariates

*The primary evaluation period for this study was the Year 2 pollen season.

[†]For ease of comparison between scores, CS and RMS results were recalculated on a 0–18 scale in this article.

[‡]CS_[0–18] was calculated as (RTSS + RMS_[0–18])/2 in this article.