Allergy 2007: 62: 1023–1028

© 2007 The Authors
Journal compilation © 2007 Blackwell Munksgaard

DOI: 10.1111/j.1398-9995.2007.01469.x

Original article

Assessment of combined symptom and medication scores for rhinoconjunctivitis immunotherapy clinical trials

Background: In randomized, double-blind, placebo-controlled clinical trials the efficacy of immunotherapy for allergic rhinoconjunctivitis is evaluated using the Average Rhinoconjunctivitis Total Symptom Score (ARTSS). Effective treatment is associated with lower ARTSS relative to placebo. For ethical reasons patients are provided with registered rescue medication, which may alleviate symptoms and thus reduce symptom scores. This effect biases the mean difference in ARTSS between effective treatment and placebo towards zero. Therefore, when rescue medication is taken by a patient, the ARTSS needs to be adjusted appropriately.
Methods: We considered five outcome measures: ARTSS, Average Rescue Medication Score (ARMS), and three combined symptom and RMSs. To assess the different outcome measures regarding their power to discriminate between effective treatment and placebo, we calculated their effect size when applied to data from a clinical trial of immunotherapy for allergic rhinoconjunctivitis.
Results: Of the five outcome measures considered, the average of the ARTSS and ARMS was associated with the largest effect size, and thus with the highest power to show treatment efficacy. Discriminant and multivariate analyses suggest that this average is approximately optimal among all weighted sums of ARTSS and ARMS.
Conclusion: Our findings support recommendations made in a World Allergy Organisation document on methodological aspects of immunotherapy trials. The average of the ARTSS and ARMS should be considered as a primary efficacy variable in clinical trials of immunotherapy for allergic rhinoconjunctivitis.

J. Clark[1], R. Schall[2]

[1]Biostatistics, Quintiles South Africa, Centurion, South Africa; [2]Biostatistics, Quintiles South Africa, Bloemfontein and Department of Mathematical Statistics, University of the Free State, Bloemfontein, South Africa

Key words: combined score; discriminant analysis; multivariate analysis; Rescue Medication Score; Rhinoconjunctivitis Total Symptom Score.

J. Clark
Biostatistics
Quintiles South Africa
Centurian, 0157
South Africa

Accepted for publication 24 May 2007

The efficacy of immunotherapy for allergic rhinoconjunctivitis is commonly evaluated using patient scores of their allergic symptoms. In randomized, double-blind, placebo-controlled clinical trials effective treatment should decrease the mean symptom scores relative to placebo. However, patients participating in clinical trials may suffer intolerable symptoms, be it because they were randomized to placebo, or because even active treatment does not alleviate all symptoms. For ethical reasons, therefore, registered rescue medication must be provided to all patients during the trial.

The relationship between symptom scores and the use of rescue medication is potentially complex. On the one hand, use of rescue medication may alleviate symptoms and thus decrease a patient's symptom score. On the other hand, if the treatment under investigation is effective, patients randomized to placebo are more likely to take rescue medication than patients randomized to active treatment, which may bias the mean difference in symptom scores between active treatment and placebo towards zero. Thus, if efficacy assessment is based on treatment differences in symptom scores alone, use of rescue medication can make it more difficult to show the superiority of active treatment over placebo. For this reason, clinical trials of the efficacy of immunotherapy for allergic rhinoconjunctivitis should always include measurements both of the symptoms and of the use of rescue medication (1–6).

When assessing the efficacy of immunotherapy for allergic rhinoconjunctivitis, symptom and Rescue Medication Scores (RMS) are often analysed independently. If this is the case, symptom scores are usually specified as the primary efficacy variable, while use of rescue medication is analysed as a secondary efficacy variable. However, because of the interdependence between symptom score and use of rescue medication, separate analysis of the respective outcome measures may not be optimal. For this reason, it has been suggested to use either some

form of adjusted symptom score (symptom score adjusted for the use of rescue medication), or a combined score consisting typically of a (weighted) sum of symptom and RMSs (1). For example, TePas et al. (2) use a combined symptom and medication score for evaluation of oral allergen immunotherapy. Furthermore, Akerlund et al. (6) state that in a clinical situation, the patient history, symptoms and signs are sufficient for a therapeutic decision. The symptom score alone, however, although relevant to the patient, may not adequately reflect the pharmacological efficacy of the specific intervention. Therefore, in a scientific context in which measurable and verifiable data are wanted there is a demand for more precise diagnostic criteria.

In the present paper, we investigate three different definitions of symptom scores adjusted for the use of rescue medication, including a proposal made in a World Allergy Organisation (WAO) statement for the sum of symptom and RMSs (1). All of these methods lead to a univariate analysis of an adjusted or combined outcome score. Furthermore, we consider the bivariate analysis of symptom and RMSs.

We propose to calculate the effect size of each outcome measure under consideration, in order to assess and compare their power to discriminate between effective treatment and placebo, and apply these methods to data from a large trial of immunotherapy for allergic rhinoconjunctivitis.

Methods

Symptom and Rescue Medication Scores

Average Rhinoconjunctivitis Total Symptom Score. A frequently used approach to score symptoms of allergic rhinoconjunctivitis is a 4-point rating scale from 0 = no symptoms to 3 = severe symptoms, applied to each symptom. The severity of symptoms is recorded by patients in a diary card, and symptoms usually include sneezing, rhinorrhoea, nasal pruritus, nasal congestion, ocular pruritus and watery eyes. The total daily score is the sum of the individual symptom scores for a given day, resulting in a rhinoconjunctivitis total symptom score (RTSS) that ranges between 0 and 18, if six symptoms are evaluated. The average RTSS (ARTSS) is the average of the daily RTSS during the period of observation, e.g. during the pollen season or during the peak pollen season. In order to make the range of this outcome measure invariant over the number of symptoms scored, we divide by the number of individual symptoms evaluated (e.g. we divided by six), so that ARTSS has a range from 0 to 3.

Average Rescue Medication Score. The registered rescue medication allowed in clinical trials must be specified in the study protocol (1), and patients should be instructed to use a stepwise regimen, stepping up as symptoms are not sufficiently alleviated, e.g. as follows:

Step 1 Antihistamine (oral or ocular formulation)
Step 2 Nasal corticosteroids
Step 3 Oral corticosteroids

Scores are assigned to the different medications as follows: 0 = no rescue medication taken; 1 = patient took antihistamines; 2 = patient took nasal corticosteroids and 3 = patient took oral corticosteroids. The RMS for a given day is equal to the highest score recorded for that day. The average RMS (ARMS) is formed as the average of the daily RMS during the period of observation. Thus ARMS has a range from 0 to 3.

Adjusted and Combined Scores

ARTSS (LOCF). The ARTSS [last observation carried forward (LOCF)] is calculated as follows: the RTSS recorded on a given day when rescue medication was used is adjusted by replacing it with the patient's RTSS recorded for the last (most recent) day before rescue medication was started (unless the replacement score is lower than the current day's score, in which case the current day's score is retained). Once a day has gone by during which the patient did not use any rescue medication, the observed score of the given day is again used as is. Effectively, therefore, the observed RTSS of the last day before rescue medication was started is carried forward until use of rescue medication has stopped again; this can be viewed as a variant of the 'LOCF' method which is frequently applied to handle missing data in clinical trials. The average adjusted RTSS is formed as the average of the total daily symptom scores adjusted in this way during the period of observation, divided by the number of symptoms. We denote this adjusted score by ARTSS (LOCF) which has a range from 0 to 3.

ARTSS (WC). The ARTSS [worst case (WC)] is calculated as follows: the RTSS recorded on a given day when rescue medication was used is adjusted by replacing it with the maximum possible score, namely 18 when six symptoms are evaluated; this can be viewed as a 'WC' method of adjustment. The average adjusted RTSS is formed as the average of the total daily symptom scores adjusted in this way during the period of observation, divided by the number of symptoms. We denote this adjusted score by ARTSS (WC) which has a range from 0 to 3.

ACS (WAO). A statement by a WAO taskforce proposes, as a pragmatic solution, to form a combined symptom and medication score as the sum of the symptom and medication scores (1). Equivalently, we form the average of the symptom and RMSs, to obtain the Average Combined Score (ACS)

$$\mathrm{ACS(WAO)} = \frac{(\mathrm{ARTSS} + \mathrm{ARMS})}{2}$$

The ACS (WAO) has a range from 0 to 3. Note that the ACS (WAO) takes into account the stepwise regimen of the three categories of rescue medication used, higher classes of rescue medication used leading to a higher ARMS and thus to a higher ACS (WAO). In contrast, the methods of adjustment described above (LOCF and 'WC') lead to adjusted scores that do not depend on the class of rescue medication used.

ACS (WS). The ACS (WAO) can be viewed as a weighted sum of the ARTSS and the ARMS, giving equal weights of 0.5 to the two summands. More generally, we might consider a Weighted Score ACS (WS) of the form

$$\mathrm{ACS(WS)} = (w_1\mathrm{ARTSS} + w_2\mathrm{ARMS}); \quad w_1 + w_2 = 1$$

Of course, the (unadjusted) ARTSS and ARMS can be viewed as special cases of ACS (WS), as ACS (WS) = ARTSS for weights $w_1 = 1$ and $w_2 = 0$ and ACS (WS) = ARMS for weights $w_1 = 0$ and $w_2 = 1$. Furthermore, as pointed out above, ACS (WS) = ACS (WAO) for weights $w_1 = 0.5$ and $w_2 = 0.5$.

When a data base of symptom and RMSs is available from a placebo-controlled clinical trial, suitable weights w_1 and w_2 can be obtained using discriminant analysis. This procedure derives a linear combination of the variables (ARTSS and ARMS) that has the highest possible multiple correlation with the two treatment groups (active *vs* placebo), that is, a linear combination is determined that best discriminates between active treatment and placebo groups.

Optimal weights for ACS (WS) determined in this way using discriminant analysis can be used to judge how close the ARTSS, ARMS and ACS (WAO) are to the optimally weighted sum score.

Data

We calculated the various outcome measures, ARTSS, ARMS, ARTSS (LOCF), ARTSS (WC) and ACS (WAO) described above for each patient in a randomized, double-blinded, placebo-controlled clinical trial of immunotherapy for allergic rhinoconjunctivitis (7). In this trial, patients were randomized to receive one of three active doses (100, 300 and 500 index of reactivity [IR] of specific sublingual immunotherapy (SLIT) or placebo. The primary efficacy variable was prospectively specified as the ARTSS during the pollen season.

The clinical trial took place during 2005 at 42 sites in 10 European countries. The pollen season in the various sites varied between 8 and 44 days, while a suitably defined peak pollen season (with particularly high pollen counts) varied between 7 and 16 days.

The statistical analysis of the trial involved analysis of covariance (ANCOVA) of the ARTSS with treatment and site as main effects, and retrospective RTSS (assessing allergic symptoms during the previous year's pollen season), age, gender, asthma (present *vs* absent) and sensitization status (mono- *vs* poly-sensitization) as covariates. Both the 300 and 500 IR doses were shown to be safe and efficacious when compared with placebo (7). The statistical assessment and comparison of the various outcome measures described in the present paper were carried out based on the data for the 300 IR dose (the dose chosen for further clinical evaluation) and placebo.

Statistical Assessment of Outcome Measures

In order to assess and compare various outcome measures regarding their power to discriminate between effective treatment and placebo, we propose to calculate their effect size, namely the estimated mean difference between active treatment (300 IR) and placebo, divided by the pooled standard deviation of the data. The greater the effect size, the greater the discriminatory power of the associated outcome measure. The *P*-value associated with the estimated mean difference between active treatment and placebo can serve a similar purpose-smaller *P*-values indicating greater discriminatory power of the associated outcome measure – but the effect size allows useful comparisons of the discriminatory power of outcome measures across data sets with different sample sizes.

Specifically, for the data set described in *Data*, active treatment and placebo were compared with respect to each of the various outcome measures using the ANCOVA model specified in *Data* [Procedure GLM of SAS (8)]. For each outcome measure, the estimated least squares mean difference between the treatments and the associated *P*-value from the ANCOVA are reported as well as the effect size.

Using discriminant analysis [Procedure CANDISC of SAS (9)], we determined the linear combination (weighted sum) of ARTSS and ARMS that best discriminated between the two treatments. We also performed a multivariate analysis of variance (MANOVA) with ARTSS and ARMS as bivariate outcome [Procedure GLM of SAS (8)].

Results

Table 1 represents summary statistics for the various outcome measures, while Table 2 represents the estimated mean differences between active treatment (300 IR) and placebo, *P*-values and effect sizes.

The difference between active treatment and placebo is statistically significant for all outcome measures considered, both when data for the peak and for the whole pollen season are considered (Table 2). However, the effect sizes associated with the various adjusted outcome measures are generally somewhat larger than the effect sizes associated with the (unadjusted) ARTSS and ARMS. This suggests that the use of adjusted scores provides somewhat greater discriminatory power between active treatment and placebo than the use of the unadjusted ARTSS (or ARMS) alone. Interestingly, the highest effect size among the five outcome measures considered is associated with ACS (WAO), the sum score (with equal weights) proposed in the WAO statement (1).

Figures 1 and 2 represent the observed ARTSS and ACS (WAO), respectively, during the peak pollen season. Figure 1 suggests that high scores of ARTSS (scores between 1.0 and 1.8) are less prevalent for active treatment than for placebo. However, the distributions of scores for active treatment and placebo, respectively, are more clearly distinguished for the combined score ACS (WAO) than for ARTSS: low scores of ACS (WAO) (between 0.0 and 0.7) are more prevalent for active treatment, and high scores (between 0.7 and 1.8) are more prevalent for placebo (Fig. 2).

The optimal linear combination of ARTSS and ARMS, and thus the largest effect size for all weighted sums of those two scores, can be found through discriminant analysis. For the data from the peak pollen season, the optimal weights are given by

$$\mathrm{ACS(WS)} = 0.45\mathrm{ARTSS} + 0.55\mathrm{ARMS}$$

Based on data from the whole pollen season, the weights are $w_1 = 0.58$ and $w_2 = 0.42$, respectively. In both cases, the optimal weights are reasonably close to 0.50, and indeed the effect size associated with the discriminant Weighted Score ACS (WS) is the same as the effect size of the simple average score ACS (WAO), when rounded to two decimal places (Table 2). This suggests that ACS (WAO) is close to optimal among all weighted sums of ARTSS and ARMS, and can serve as a good – and at the same time pragmatic – approximation of the optimal score.

Table 1. Outcome measures for rhinoconjunctivitis clinical trial data

Pollen season	Treatment	Statistic	ARTSS	ARMS	ARTSS (LOCF)	ARTSS (WC)	ACS (WAO)	ACS (WS)
Peak season	300 IR	*n*	121	121	121	121	121	121
		Mean	0.60	0.29	0.64	1.02	0.45	0.43
		SD	0.56	0.39	0.59	0.83	0.40	0.39
		Min.	0.00	0.00	0.00	0.00	0.00	0.00
		Q1	0.17	0.00	0.17	0.32	0.16	0.14
		Median	0.47	0.14	0.50	0.88	0.36	0.35
		Q3	0.87	0.50	0.94	1.56	0.64	0.63
		Max.	2.65	2.00	2.80	3.00	1.89	1.87
	Placebo	*n*	135	135	135	135	135	135
		Mean	0.80	0.47	0.88	1.38	0.63	0.62
		SD	0.57	0.58	0.64	0.97	0.47	0.47
		Min.	0.00	0.00	0.00	0.00	0.00	0.00
		Q1	0.31	0.00	0.34	0.52	0.24	0.22
		Median	0.68	0.22	0.72	1.33	0.57	0.56
		Q3	1.27	0.89	1.36	2.17	0.97	0.96
		Max.	2.40	2.00	2.60	3.00	1.88	1.89
Whole season	300 IR	*n*	123	123	123	123	123	123
		Mean	0.58	0.30	0.65	1.00	0.44	0.46
		SD	0.48	0.41	0.54	0.80	0.37	0.38
		Min.	0.00	0.00	0.00	0.00	0.00	0.00
		Q1	0.21	0.00	0.22	0.35	0.16	0.17
		Median	0.48	0.13	0.53	0.81	0.36	0.39
		Q3	0.84	0.43	0.92	1.44	0.68	0.68
		Max.	2.15	2.00	2.57	3.00	2.01	2.03
	Placebo	*n*	136	136	136	136	136	136
		Mean	0.80	0.47	0.91	1.35	0.63	0.66
		SD	0.54	0.52	0.62	0.86	0.43	0.43
		Min.	0.00	0.00	0.00	0.00	0.00	0.00
		Q1	0.32	0.00	0.41	0.62	0.28	0.31
		Median	0.71	0.30	0.78	1.22	0.55	0.60
		Q3	1.23	0.77	1.38	2.03	0.87	0.93
		Max.	2.36	2.00	2.55	3.00	1.81	1.77

ARTSS, Average Rhinoconjunctivitis Total Symptom Score; ARMS, Average Rescue Medication Score; ARTSS (LOCF), ARTSS with last observation carried forward adjustment for rescue medication usage; ARTSS (WC), ARTSS with worst case adjustment for rescue medication usage; ACS (WAO), average combined ARTSS and ARMS as proposed by World Allergy Organisation; ACS (WS), discriminant Weighted Score; *n,* number of patients; SD, standard deviation; Min., minimum; Q1, first quartile 1; Q3, third quartile; Max., maximum.

Table 2. Effect size of outcome measures

Pollen season	Treatment contrast	Statistic	ARTSS	ARMS	ARTSS (LOCF)	ARTSS (WC)	ACS (WAO)	ACS (WS)
Peak season	300 IR *vs* Placebo	Effect size	−0.34	−0.35	−0.37	−0.39	−0.42	−0.42
		LS mean difference	−0.19	−0.18	−0.23	−0.36	−0.18	−0.18
		P-value	0.0048	0.0044	0.0020	0.0013	0.0006	0.0029
Whole season	300 IR *vs* Placebo	Effect size	−0.43	−0.35	−0.46	−0.44	−0.48	−0.48
		LS mean difference	−0.22	−0.17	−0.27	−0.36	−0.19	−0.20
		P-value	0.0004	0.0029	0.0001	0.0003	<0.0001	0.0004

ARTSS, Average Rhinoconjunctivitis Total Symptom Score; ARMS, Average Rescue Medication Score; ARTSS (LOCF), ARTSS with last observation carried forward adjustment for rescue medication usage; ARTSS (WC), ARTSS with worst case adjustment for rescue medication usage; ACS (WAO), average combined ARTSS and ARMS as proposed by World Allergy Organisation; ACS (WS), discriminant Weighted Score; LS mean difference, least squares mean difference from ANCOVA; Effect size, LS mean difference/pooled standard deviation from ANCOVA.

P-values are from the univariate ANCOVA for the respective outcome measure; *P*-value for ACS (WS) from MANOVA.

We note that the discriminant Weighted Score ACS (WS) has weights determined from the data (the weights are themselves random variables) so that a *P*-value for treatment obtained from a univariate ANCOVA of the discriminant score would not be appropriate. A valid analysis is through MANOVA with the bivariate outcome ARTSS and ARMS. Using this analysis, the *P*-values for treatment for the peak and for the whole pollen season

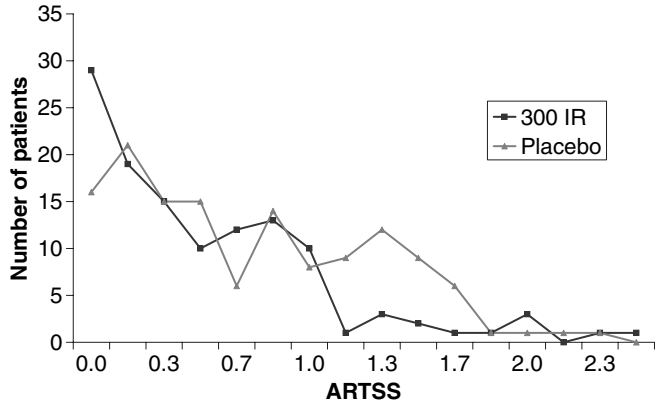


Figure 1. Line graph of Average Rhinoconjunctivitis Total Symptom Score (ARTSS) during the peak pollen season.

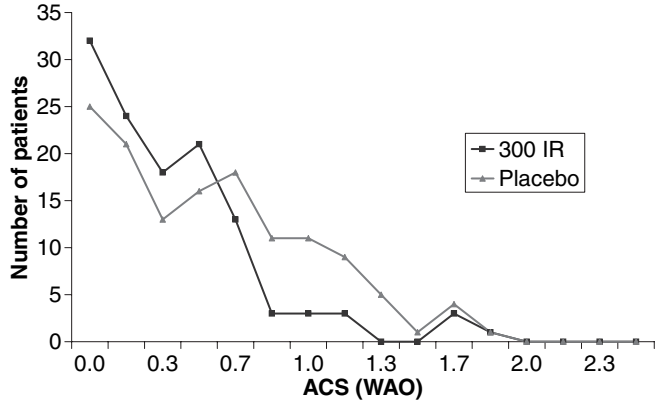


Figure 2. Line graph of average combined score (ACS [WAO]) during the peak pollen season.

are 0.0029 and 0.0004, respectively. Indeed, discriminant and multivariate analysis are closely related: the characteristic vector associated with the MANOVA, without covariates, is equal to the ACS (WS) (10).

Discussion

Effective immunotherapy of allergic rhinoconjunctivitis should decrease both the incidence and severity of allergic symptoms, and the use of concomitant rescue medication. For this reason, both symptom scores and use of rescue medication must be recorded as outcome measures in clinical trials of immunotherapy of allergic rhinoconjunctivitis (1). These two outcome measures are interdependent and their relationship is complex, and potentially many methods for adjusting the symptom score for use of rescue medication can be thought of. In the present paper, we have considered five outcome measures: ARTSS; ARMS; two methods of adjustment of ARTSS, namely LOCF and 'WC' and finally, a method for combining ARTSS and ARMS by forming the average of the two (equivalently: the sum of the two) as proposed in a recent WAO statement (1).

The discriminatory power of those outcome measures, and in fact of other outcome measures or scoring methods that might be proposed, can be assessed by calculating their effect size. The results of our study suggest that both outcome measures, ARTSS and ARMS, provide information on treatment efficacy, and in fact, that outcome measures based on the combination of symptom and RMSs provide better discriminatory power than each individual outcome measure, ARTSS or RMS, alone.

It turns out that of the five outcome measures considered by us, the average of the ARTSS and RMS is associated with the largest effect size, and thus with the highest power to show treatment efficacy. Discriminant and multivariate analysis suggest that this average score is approximately optimal among all weighted sums of ARTSS and RMS.

When clinical trials of immunotherapy for allergic rhinoconjunctivitis are planned, data bases from similar clinical trials are usually available (e.g. prior phase II or III clinical trials with the same or similar treatment). If this is case, our methodology can be used to assess the discriminatory power of various outcome measures considered for the planned clinical trial.

We may note that the statistical precision of an outcome measure is not necessarily the same as its clinical relevance. However, when selecting a particular outcome measure for a clinical trial, the measure with the highest precision and discriminatory power should be used, providing it is relevant and of clinical interest to the question investigated. As pointed out by Canonica et al. (1), immunotherapy of allergic rhinoconjunctivitis potentially has an effect both on rhinoconjunctivitis symptom scores, and on use of rescue medication, thus a combination of these two outcome measures may provide clinically more relevant information than each of these two outcome variables alone. The combined score, therefore, seems to be both a clinically more relevant, and, as our results suggest, a statistically a more precise outcome measure for immunotherapy trials than either symptom scores or RMSs alone.

Our results on the relative merits of the various outcome measures are based on applying those outcome measures to data from a large clinical trial of immunotherapy for allergic rhinoconjunctivitis. Our findings provide preliminary support for the recommendation made in a WAO statement (1) on methodological aspects of immunotherapy trials, namely that the average (equivalently: the sum) of the ARTSS and ARMS should be considered as the primary efficacy variable in clinical trials of immunotherapy for allergic rhinoconjunctivitis.

References

1. Canonica GW, Baena-Cagnani CE, Bousquet J, Bousquet PJ, Lockey RF, Malling HJ 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**:317–324.
2. TePas EC, Hoyte EG, McIntire JJ, Umetsu DT. Clinical efficacy of microencapsulated timothy grass pollen extract in grass-allergic individuals. Ann Allergy Asthma Immunol 2004;**92**: 25–31.
3. Dahl R, Kapp A, Colombo G, de Monchy JGR, Rak S, Emminger W et al. Efficacy and safety of sublingual immunotherapy with grass allergen tablets for seasonal allergic rhinoconjunctivitis. J Allergy Clin Immunol 2006a;**118**:434–440.
4. Dahl R, Stender A, Rak S. Specific immunotherapy with SQ standardized grass allergen tablets in asthmatics with rhinoconjunctivitis. Allergy 2006b;**61**:185–190.
5. Durham SR, Yang WH, Pedersen MR, Johansen N, Rak S. Sublingual immunotherapy with once-daily grass allergen tablets: a randomized controlled trial in seasonal allergic rhinoconjunctivitis. J Allergy Clin Immunol 2006;**117**:802–809.
6. Akerlund A, Andersson M, Leflein J, Lildholdt T, Mygind N. Clinical trial design, nasal allergen challenge models, and considerations of relevance to pediatrics, nasal polyposis, and different classes of medication. J Allergy Clin Immunol 2005;**115**:S460–S482.
7. Didier A, Melac M, Combebias A, Andre C. Efficacy and safety of sublingual immunotherapy (SLIT) tablets in grass pollen rhinoconjunctivitis. J Allergy Clin Immunol 2006;**117**:721 (Abstract).
8. SAS Institute Inc SAS/STAT 9.1 User's Guide, Volume 1. The GLM Procedure. Cary, NC, USA: SAS Institute Inc., 2004:1731–1906.
9. SAS Institute Inc. SAS/STAT 9.1 User's Guide, Volume 1. The CANDISC Procedure. Cary, NC, USA: SAS Institute Inc., 2004:781–809.
10. Muirhead RJ. Aspects of multivariate statistical theory. New York: Wiley, 1982.