

Review article

Sublingual immunotherapy: efficacy – methodology and outcome of clinical trials

Using Medline we identified 39 placebo-controlled, double-blind sublingual immunotherapy (SLIT) studies providing symptom–medication score. These were retrospectively evaluated for evidence of clinical efficacy and quality of study presentation. Clinical efficacy was estimated according to statistical significance and graded as: *unequivocal efficacy* (statistically significant difference from placebo in both symptom and medication scores or the combined score), which was observed in 28% of studies, *possible efficacy* (significant improvement in either symptom or medication scores) seen in 33%, and *no efficacy* (no statistical difference between active treatment and the placebo group), as found in 38% of studies. Generally, studies were limited by the number of patients, showed a high frequency of withdrawals, a short duration of treatment, and insufficient data on randomization. The magnitude of efficacy additional to placebo treatment must be > 20% in order to justify the treatment. This review concludes that future SLIT studies should be planned in accordance with international recommendations in order to be conclusive.

H.-J. Malling

Allergy Clinic, National University Hospital,
Copenhagen, Denmark

Key words: clinical efficacy; clinical trial; quality criteria; sublingual immunotherapy.

Professor Hans-Jørgen Malling
Allergy Clinic
National University Hospital
Copenhagen
Denmark

Introduction

It is only for the past 50 years that the clinical efficacy of allergen-specific immunotherapy has been assessed with controlled studies. In the evidence-based world of modern medicine, however, all treatments developed for patients must at least demonstrate unquestionable clinical efficacy; in other words, there should be solid scientific documentation of the benefit from intervention (1).

In to this evidence-based climate, this review provides a critical survey of 39 placebo-controlled, double-blind studies of sublingual immunotherapy (SLIT), published in English-speaking journals between 1990 and March 2006. These studies form the basis of the evidence for the clinical efficacy of SLIT. The studies were evaluated exclusively on the available printed information, and consequently the quality of the underlying experimental design could not be assessed. All studies were categorized in relation to the quality of information presented by the authors following the criteria described in this review and were further analysed to determine clinical efficacy.

Estimation of clinical efficacy

Clinical efficacy was categorized as *unequivocal efficacy* (statistically significant difference from placebo in both symptom and medication score, or the combined score), *possible efficacy* (significant improvement in either

symptom or medication scores), and *no efficacy* (no statistical difference between active treatment and the placebo group). The magnitude of efficacy was estimated as the percentage reduction in the combined symptom–medication score (assuming equivalent importance of symptoms and medication scores) in the active vs the placebo group (2). The magnitude of efficacy was calculated as described by Malling (2). The complete list of studies included in this review is available from the author.

Quality criteria for conclusive studies

Strict criteria for performing scientific studies are essential to generate data for clinical efficacy of appropriate modern standards, to ensure that patients are offered treatment based on scientific evidence, and to minimize the spending of limited financial resources on badly planned studies. To this end, several journals only accept studies conducted according to the Consort guidelines (3). These guidelines provide criteria for the evaluation of the quality (and consequently the reliability) of published studies but, perhaps more importantly, give help for planning conclusive studies. Furthermore, the World Allergy Organization has formed a working group to define the quality criteria for immunotherapy studies, which should provide guidance in improving the standards and quality of immunotherapy studies (Allergy In press).

Analysis of the published literature

This review only includes placebo-controlled and double-blind studies. The number of patients screened, randomized and completing a study is essential information for transferring conclusions from a clinical trial to daily clinical practice. Papers were analysed for information on the number of participants randomized, receiving treatment, completing the study protocol, and outcomes. The duration of treatment was also assessed. The evaluation of the 39 clinical trials considered the following:

- 1 *Randomization*, the allocation of patients to treatment groups is essential for ensuring the balance between groups. The 39 papers were evaluated for the details given about the randomization process.
- 2 *Pretreatment monitoring*, was patient inclusion based on a defined period of pretreatment monitoring or on memory recall of disease severity?
- 3 *Blinding of intervention and assessment*, does the paper describe how interventions and assessments were blinded and how successful the blinding was?
- 4 *Definition of outcomes*, were primary and secondary outcomes clearly defined?
- 5 *Participant flow*, to evaluate efficacy the flow of participants through each stage of the study should be described carefully.
- 6 *Handling of missing data*, is the missing data described in the publication?
- 7 *Estimation of efficacy*, was the efficacy of active treatment evaluated in comparison with the placebo group? Was it based on intergroup or intragroup analysis?
- 8 *Statistical methods*, are data on applied statistics available and why were these tests used?

Eleven of the 39 studies (28%) were categorized as showing 'unequivocal efficacy' in 243 actively treated patients, 13 (33%) studies demonstrated 'possible efficacy' (361 patients) and 'no efficacy' was observed in 15 studies based on 414 patients (38%). The magnitude of efficacy compared with improvement obtained in the placebo group is shown in Fig. 1.

The studies could be grouped by the number of participants: two studies involved <10 patients; 18 considered 10–20 patients; 15 included 21–50 patients and only four investigated more than 50 patients. Withdrawal rates of included patients of <5% was observed in 13 studies, in 10 rates of 6–10% were noted, in a further 13 trials it was 10–25% and in five trials more than 25% of patients withdrew. Treatment was for <6 months in 14 of the studies (36%), between 6 and 12 months in 10 trials (27%) and for more than 1 year in 15 studies.

Detailed information on the process of randomization was given in only 23% (9/39) of studies. Allocation to treatment groups based on pretreatment monitoring

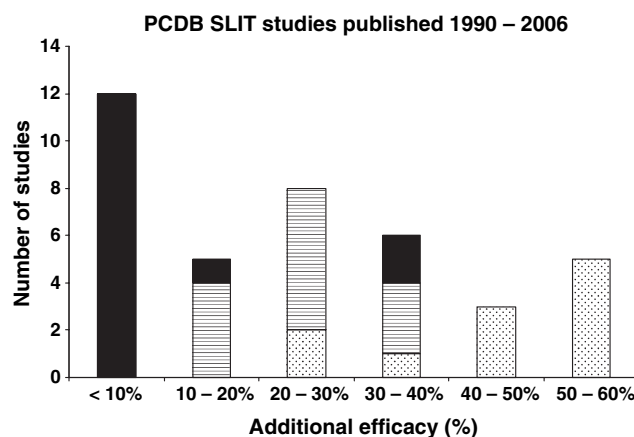


Figure 1. The magnitude of efficacy (additional to the placebo effect) given in 10% intervals. Solid bars represent 'no efficacy'; hatched bars, 'possible efficacy'; and dotted bars, 'efficacy'.

was performed in 36% (14/39) of studies. The success of blinding of intervention was poorly described, only two of 39 (5%) papers included this information. Six studies (15%) failed to define primary outcome measures, they were partly described in nine (23%), and well defined in 24 (62%) of the studies. The flow of participants including patients leaving the study was described in 46% (18/39) of studies. Information on the handling of missing data was only given in four (10%) studies.

Most studies, 34 of 39 (87%), presented data on intergroup analysis (efficacy of active treatment evaluated in comparison with placebo). The statistical analysis was parametric in five (13%) studies and nonparametric in 34 (87%). The size of treatment effect as estimated by 95% confidence intervals or equivalent was stated in 20 (51%) studies.

Quality of sublingual studies

Most studies included too few patients to ensure a high power of detecting significant changes in disease severity. When planning a study, the number of patients should be large enough to have high probability (power) to detect statistical significance. The size of the clinical effect to be evaluated is inversely related to the sample size necessary to detect it (large samples are necessary to detect small differences) (4). Inclusion of insufficient numbers of patients increases the risk of statistical type I errors, and can also cause type II statistical errors if there are differences in disease severity between the evaluated groups at inclusion. A major problem in many studies is the high frequency of withdrawals, which reduces the applicability to daily clinical practice. The duration of treatment has to be sufficient for the immunological effects of immunotherapy to take place.

In the optimal clinical study, participants should be assigned to comparable groups on the basis of a random process, which minimizes the risk of bias in group assignment in terms of size and characteristics (5–7). Randomization is extremely important to achieve balance between groups in clinical trials with small numbers of patients.

Pretreatment monitoring is often not included in clinical studies because it is time-consuming and expensive. It is however, useful in allocating patients to treatment groups and, consequently, increases the likelihood of achieving clinical efficacy (8). Direct evaluation of data obtained from placebo and actively treated groups is only useful if comparing similar groups with respect to parameters important for the primary outcome (9). A pretreatment monitoring period allows determination of disease severity based on objective parameters and patients can be randomized appropriately, without bias. Bias can also be avoided by blinding of the intervention. The ‘true’ treatment effect associated with the intervention can be evaluated by combining the use of placebo with blinding. The success of blinding should be evaluated and reported (10).

The primary outcome of a SLIT study should be the combined symptom–medication score. None of the 39 studies evaluated the scoring system used, probably because no universal scoring system exists for either clinical symptoms or the use of rescue drugs. The primary outcome must be related to the disease treated.

The reliability of a study becomes questionable if information is missing on treatment compliance, and if patients are lost to follow-up or excluded from the analysis. Ideally, the primary outcome (clinical efficacy) should be based on all patients randomized (intent-to-treat analysis). Participants leaving the study due to insufficient clinical efficacy must be included in the primary outcome analysis to evaluate the ‘true’ treatment effect. Likewise, patients leaving the study due to adverse effects should be included in the safety analysis.

Randomized, controlled trials aim to compare groups of participants that differ only with respect to the intervention (intergroup analysis). Study results should be reported for each outcome as the results in each group and the difference between the groups. This gives the parameter known as the magnitude of treatment effect (Table 1).

The choice between parametric and nonparametric statistics should be discussed. In biological science most small samples are not normally distributed. This may be overcome partly by transforming data, but showing statistically significant difference using nonparametric tests usually reinforces the conclusion. The interpretation of clinical study results can be improved by the inclusion of confidence intervals that give the range of the differences or the magnitude of efficacy between two values, in other words, the range of uncertainty, which is expected to include the ‘true’ value.

Clinical efficacy

Almost 40% of the published studies of SLIT showed no statistically significant difference between active and placebo treatment. Unequivocal efficacy (significant difference in both symptom and medication score or combined score) was observed in only 28% of studies. It could be argued that the category ‘possible efficacy’ (significant difference in either symptom or medication scores) represents a true treatment effect. Taking this approach, 60% of these SLIT studies show a positive treatment effect, which is in agreement with other reviews (9, 11, 12).

Minimal clinically relevant efficacy

In the discussion of clinical efficacy it is important to remember that allergic patients suffer from symptoms controlled with anti-allergic drugs. Consequently, efficacy should be based on a reduction in these parameters. Statistically significant but clinically irrelevant differences may be observed by including high numbers of participants. The magnitude of efficacy should be clinically relevant, in other words, the reduction in symptom scores and drug consumption should significantly reduce the morbidity of the disease. An attempt to define the minimal clinically relevant reduction in disease severity in subcutaneous immunotherapy has been published (2). For SLIT to have a better performance than antihistamine treatment the magnitude of efficacy should be at least 20% more than the effect of placebo. The present review shows that even when the treatment effect is >30%, some studies (14%) are negative in this respect.

Conclusion

In this review, analysis of published placebo-controlled, double-blind clinical trials of SLIT showed that, depending on the definition of clinical efficacy used, 40–72% of published studies of SLIT clinical trials failed to show efficacy of the active treatment. This does not necessarily reflect the lack of efficacy of SLIT *per se*, but may reflect a deficiency in clinical trial design, number and type of patients recruited, and statistical analysis or reportage. Therefore, to achieve improved assessment of SLIT trials, it is recommended that future SLIT studies should be planned accordingly:

- Studies should be placebo-controlled, double-blind and randomized.
- The sample size should be large enough to detect statistically significant and clinical relevant changes in disease severity.

Table 1. Quality evaluation of placebo-controlled, double-blind SLIT papers

Study no.	Public. year	Sample size*	Percentage completed (%)	Duration of treatment (months)	Randomization†	Pretreatment monitoring	Blinding of intervention‡	Definition of outcomes§	Patient flow¶	Missing data**	Intergroup analysis††	Statistical methods‡‡	Clinical efficacy§§
1	1990	66	88	18	No info	No	No	No	No	No	No	NP (No)	Possible
2	1993	44	93	3 3/4	No info	No	Yes	Partly	Yes	Yes	Yes	NP (No)	No efficacy
3	1994	15	100	2	No info	No	No	Partly	No	No	Yes	NP (No)	No efficacy
4	1994	58	100	4 1/4	No info	No	No	Partly	No	No	Yes	NP (No)	Efficacy
5	1995	34	100	3 1/2	No info	No	No	No	No	No	Yes	NP (No)	Efficacy
6	1995	31	100	10	No info	No	No	Partly	No	No	Yes	NP (No)	Possible
7	1997	30	100	12	No info	No	No	No	No	No	Yes	NP (No)	No efficacy
8	1998	20	95	24	No info	Yes	No	Yes	Yes	No	Yes	NP (Yes)	Possible
9	1998	136	88	7	No info	No	No	Yes	No	No	Yes	NP (Yes)	Possible
10	1998	41	83	4	No info	No	No	Yes	No	No	Yes	NP (No)	No efficacy
11	1998	30	93	12	No info	No	No	No	No	No	Yes	P (No)	Efficacy
12	1998	66	97	2 × 7	No info	Yes	No	Yes	No	No	Yes	NP (Yes)	No efficacy
13	1998	57	93	10	No info	No	No	Yes	No	No	Yes	NP (Yes)	Possible
14	1999	85	59	25	No info	Yes	No	Partly	Yes	No	Yes	NP (Yes)	No efficacy
15	1999	26	100	12	No info	Yes	No	Yes	No	No	No	NP (No)	Efficacy
16	1999	41	80	24	No info	No	No	Yes	No	No	Yes	NP (Yes)	Possible
17	1999	126	98	4	No info	No	No	Yes	Yes	No	Yes	NP (Yes)	No efficacy
18	1999	30	100	9	No info	No	No	Yes	No	Yes	Yes	NP (No)	Efficacy
19	1999	30	90	5	No info	Yes	No	Yes	Yes	No	No	NP (No)	Efficacy
20	2000	72	54	24	No info	No	No	Partly	Yes	No	Yes	NP (Yes)	No efficacy
21	2000	24	87	24	Yes	Yes	No	Yes	Yes	No	Yes	NP (No)	Efficacy
22	2000	48	92	3 1/2	Yes	No	No	Partly	Yes	No	Yes	NP (Yes)	Possible
23	2001	15	100	6	No info	Yes	No	Yes	No	No	Yes	NP (No)	No efficacy
24	2001	30	90	5	No info	Yes	No	Yes	Yes	No	Yes	NP (No)	No efficacy
25	2001	20	100	12	No info	No	No	Yes	No	No	Yes	NP (Yes)	Efficacy
26	2002	56	88	18	Yes	No	No	Yes	No	Yes	Yes	NP (Yes)	No efficacy
27	2003	60	75	24	No info	No	No	Partly CPT	Yes	No	Partly only sympt.	NP (No)	Possible
28	2003	106	87	7 1/2	No info	No	No	Yes	No	No	Yes	P (Yes)	Possible
29	2003	28	79	24	Yes	No	No	Yes	Partly	No	Yes	NP (No)	Possible
30	2003	30	90	13	Yes	No	No	Partly	No	No	Yes	NP (No)	No efficacy
31	2003	86	100	6	No info	Yes	No	No	No	No	No	NP (Yes)	Possible
32	2004	47	79	22	Yes	Yes	Yes	Yes	Yes	Yes	Yes	NP (Yes)	Efficacy
33	2004	32	56	24	No info	Yes	No	No	Yes	No	Yes	NP (Yes)	Possible
34	2004	161	91	12	Yes	Yes	No	Yes	No	No	Yes	NP (Yes)	No efficacy
35	2004	120	74	24	Yes	Yes	No	Yes	No	No	Yes	P (No)	Possible
36	2004	83	67	2 1/2	No info	No	No	Yes	Yes	No	Yes	NP (Yes)	No efficacy
37	2004	97	77	32	Yes	No	No	Yes	Yes	No	Yes	NP (Yes)	Efficacy
38	2006	114	82	5 1/2	No info	No	No	Yes	Yes	Yes	Yes	P (Yes)	Efficacy
39	2006	110	88	6	No info	Yes	No	Yes	Yes	No	Yes	P (Yes)	No efficacy

*Total number of patients included.

†Allocation to intervention groups based on a random process incl. description of criteria for randomization.

‡The success of the blinding evaluated.

§Detailed information on the primary and secondary outcomes and how they were measured.

¶Information for each group on the number of participants randomized, receiving treatment, completing the study protocol, and analysed for outcomes.

**Handling of missing data described.

††Efficacy of active treatment evaluated in comparison with placebo (intergroup analysis).

‡‡Parametric or nonparametric statistics applied (95% confidence interval, SD, SEM considered).

§§Clinical efficacy scores as efficacy, possible efficacy and no efficacy.

Malling

- The duration of treatment should be sufficient to allow clinical improvement to occur.
- Patients should be selected according to predefined clinical criteria.
- The primary and secondary outcome measures should be clearly defined.
- The participant flow should be described.

References

1. Guyatt GH, Sackett DL, Sinclair JC, Hayward R, Cook DJ, Cook RJ. Users' guides to the medical literature. IX. A method for grading health care recommendations. Evidence-Based Medicine Working Group. *JAMA* 1995;**274**:1800–1804.
2. Malling HJ. Immunotherapy as an effective tool in allergy treatment. *Allergy* 1998;**53**:461–472.
3. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *Ann Intern Med* 2001;**134**:657–662.
4. Campbell MJ, Julious SA, Altman DG. Estimating sample sizes for binary, ordered categorical, and continuous outcomes in two group comparisons. *BMJ* 1995;**311**:1145–1148.
5. Altman DG, Bland JM. How to randomise. *BMJ* 1999;**319**:703–704.
6. Enas GG, Enas NH, Spradin CT, Wilson MG, Wiltse CG. Baseline comparability in clinical trials. *Drug Inf J* 1990;**24**:541–548.
7. Treasure T, MacRae KD. Minimisation: the platinum standard for trials? Randomisation doesn't guarantee similarity of groups; minimisation does. *BMJ* 1998;**317**:362–363.
8. Frew AJ, White PJ, Smith HE. Sublingual immunotherapy. *J Allergy Clin Immunol* 1999;**104**:267–270.
9. Malling HJ. Is sublingual immunotherapy clinically effective? *Curr Opin Allergy Clin Immunol* 2002;**2**:523–531.
10. Devereaux PJ, Manns BJ, Ghali WA, Quan H, Lacchetti C, Montori VM et al. Physician interpretations and textbook definitions of blinding terminology in randomized controlled trials. *JAMA* 2001;**285**:2000–2003.
11. Canonica GW, Passalacqua G. Noninjection routes for immunotherapy. *J Allergy Clin Immunol* 2003;**111**:437–448.
12. Wilson DR, Lima MT, Durham SR. Sublingual immunotherapy for allergic rhinitis: systematic review and meta-analysis. *Allergy* 2005;**60**:4–12.