

## ***Primary and Secondary Endpoints in Clinical Trials***

Lars Jacobsen

### ***Background***

The efficacy and safety of therapeutic allergen products must be demonstrated by clinical trials that follow the guidance in “Good Clinical Practice: Consolidated Guideline” (ICH E6) adopted by the ICH, May 1, 1996. The role of statistics in clinical trial design and analysis is acknowledged as essential in that ICH guideline and described in detail in “Note for guidance on statistical principles for clinical trials” [1]. The principles outlined in the statistical guidance are primarily relevant to clinical trials conducted in the later phases of development, many of which are confirmatory trials of efficacy. In addition to efficacy, confirmatory trials may have as their primary variables a safety variable, a clinical laboratory variable or a pharmacodynamic or a pharmacokinetic variable.

In any case it is important to consider the robustness of data and study performance which is the basis for judgment of the treatment effect and primary conclusions of the trial. The interpretation of statistical measures of the treatment effect and treatment comparisons should involve consideration of the potential contribution of bias to the p-value, confidence interval, or inference.

The clinical development plan for a potential new drug is aiming to identify a dose range and schedule at which the drug can be shown to be simultaneously safe and effective, to the extent that the risk-benefit relationship is acceptable. Also the particular subjects, who may benefit from the drug, and the specific indications for its use, need to be defined.

Defining the aim of the different trials in a development plan also defines the opportunities of design, degrees of freedom for flexibility and to which extend the data can be used for either defining new hypotheses or make final conclusions.

## **Confirmative and explorative trials**

There are two main categories of clinical trials, being either confirmative or explorative. Confirmatory trials are needed to provide firm evidence of efficacy or safety. In such trials the key hypothesis of interest follows directly from the trial's primary objective, is always pre-defined, and is the hypothesis that is subsequently tested when the trial is complete. Confirmatory trials are usually based on earlier clinical work carried out in a series of exploratory studies.

*The characteristics for confirmative trials are the following.*

- Confirmatory trials should be adequately controlled.
- Are needed to provide firm evidence of efficacy or safety.
- Hypothesis is always pre-defined.
- They provide firm evidence in support of claims – adherence to protocols and SOPs.
- The principal features of the planned analysis should be set out in the protocol.
- The study addresses a limited number of questions.
- It should answer each key clinical question relevant to the efficacy or safety claim.
- Possible generalization to the intended patient population should be explained.
- The results of the confirmatory trials should be robust.
- In some circumstances evidence from a single confirmatory trial may be sufficient proof of efficacy and safety.

Explorative clinical trials should have defined and precise objectives. Their objectives may not always lead to simple tests of pre-defined hypotheses and exploratory trials may imply a flexible approach to the design so that changes can be made in response to accumulating results. Tests of hypothesis may be carried out, but the choice of hypothesis may be data dependent. Exploratory trials cannot be used for final proof of efficacy, although they may contribute to the total body of data and information.

*The characteristics for exploratory trials are the following.*

- Exploratory studies are the platform for confirmatory trials.
- Studies should have clear and precise objectives.
- Objectives may not always lead to simple tests of pre-defined hypotheses.
- Allows flexibility design – changes can be made in response to accumulating results.
- Analysis includes data exploration.
- Hypothesis may be data dependent.
- Cannot be used as formal proof of efficacy.
- Contribute to the total body of relevant evidence.

Some trials may have both confirmatory and exploratory aspects. In confirmatory trials data are often subjected to exploratory analyses serving as a basis for further hypotheses for later research. Protocols should make a clear distinction between the aspects of a trial which will be used for confirmatory proof and the aspects which will provide data for exploratory analysis.

## ***Dose-response trials***

An important issue related to the clinical development plan is to establish the dose-response characteristics of treatments. Dose-response trials may serve a number of objectives, amongst which some are of particular interest and importance. The dose-response design can be applied to confirmation of efficacy and safety as well as investigation of the shape and location of the dose-response curve. Based on dose-response studies an appropriate starting dose can be identified and strategies for individual dose adjustments can be defined. Of great importance is the determination of a maximal dose beyond which additional benefit would be unlikely to occur.

## ***Statistical variables***

In any case it is of crucial importance for the design of the clinical development plan and mainly for statistical reliability reasons to decide on one and only one most important outcome parameter from each clinical trial. This parameter should be the primary variable.

## ***Primary variable***

The primary variable should be the variable capable of providing the most clinically relevant and convincing evidence directly related to the primary objective of the trial.

*The characteristics for the primary endpoint variable.*

- The primary endpoint will usually be an efficacy variable but safety/tolerability may sometimes be defined as the primary variable.
- Quality of life and health economics are potential primary variables.
- The selection of the primary variable should reflect norms and standards.
- The use of a reliable and validated variable is recommended.
- The primary variable should be the one used when estimating the sample size.
- The protocol should define the primary variable as it will be used in the statistical analysis.
- The primary variable and the rationale for its selection should be specified in the protocol.
- Redefinition of the primary variable after unblinding is usually unacceptable.
- When the clinical effect is to be measured in more than one way, the protocol should identify one of the measurements as the primary variable.

It may sometimes be desirable to use more than one primary variable, each of which could be sufficient to cover the range of effects of the therapies. For clinical trials on allergen-specific immunotherapy this could be the use of symptom score as well as symptom reducing medication score. The extent of intercorrelation among the proposed primary variables may be considered. It should be made clear how each of these defined variables, would be considered necessary to achieve the trial objectives. The primary hypothesis of interest should be clearly stated and the approach to statistical inference described. A justified hierarchy for the statistical

analysis should be defined and justified in the sense that only if variable 1 is statistically significant it can be justified to continue testing variable 2.

In the case when a single primary variable cannot be selected from multiple measurements associated with the primary objective, another useful strategy is to integrate or combine the multiple measurements into a single or ‘composite’ variable, using a pre-defined algorithm. In this case, the primary variable will arise as a combination of multiple clinical measurements (combined symptom and medication score). The method of combining the multiple measurements should be specified in the protocol, and an interpretation of the resulting scale should be provided in terms of the size of a clinically relevant benefit.

### ***Secondary variables***

Secondary variables are also of great importance but the outcome of variables defined as secondary variables can never be used as final documentation of efficacy. The importance of secondary parameters is underlined by the importance of data exploration mainly for the purpose of generating a new hypothesis.

*The characteristics for the secondary endpoint variable.*

- Secondary variables are either supportive measurements related to the primary objective or measurements of effects related to the secondary objectives.
- Their pre-definition in the protocol is important, as well as an explanation of their relative importance and roles in interpretation of trial results.
- The number of secondary variables should be limited and should be related to the limited number of questions to be answered in the trial.
- Paraclinical variables are secondary endpoints

### ***Indirect and paraclinical variables***

Other parameters used can be defined as surrogate variables. These are defined as indirect variables used when direct assessment of the clinical benefit is not practical. None of these surrogate markers or paraclinical parameters (objective parameters) can replace the primary clinical outcomes but can be assessed as secondary outcomes in order to investigate particular aspects of SIT treatments, e.g. mechanism of action. The main concerns are that they may not predict a clinical outcome of interest. It may measure treatment activity associated with one specific pharmacological mechanism and it may not provide full information on the range of actions. The value of the evidence for paraclinical parameters depends upon the biological relationship the demonstration in epidemiological studies of the prognostic value of the surrogate for the clinical outcome.

### ***Variables in allergen immunotherapy trials***

Recently the World Allergy Organization has proposed a model for standardization of clinical trials of allergen-specific immunotherapy for respiratory allergic symptoms [2]. The taskforce suggests defining the primary outcome of immuno-

therapy trials as the total clinical score, calculated as the sum of symptom score plus drugs usage calculated as one score. It is further stated that clinical trials of efficacy always must include a measurement of the symptoms and the use of concomitant medications, which represent the primary outcome, and that patient-rated scores are preferred as a primary outcome. The score systems used should be validated.

The secondary outcomes are defined as

- Scores of individual symptoms separately and changes in drug usage
- Quality of Life assessment
- Safety (total number of adverse events, and SIT-related side effects)
- Functional measures (e.g. lung function, rhinometry)
- Cost-effectiveness (e.g. QALY)

Surrogate markers and objective parameters are defined as

- Allergen specific reactivity (sensitivity)
  - ▶ Skin: end point skin test, late cutaneous response
  - ▶ Nose, eye and bronchi: allergen specific provocation test
- Nonspecific organ reactivity
  - ▶ Bronchial challenge with methacholine, carbachol, histamine, AMP
- Immunological parameters
  - ▶ Total and allergen specific IgE and IgG subclasses
  - ▶ Mucosal IgA
  - ▶ Lymphocyte subsets and cytokines (e.g. IL-12, IFNs, IL-5, IL-10)
  - ▶ Local and systemic inflammatory markers (e.g. adhesion molecules, urinary leukotrienes, sECP, tryptase)

The recommendations from WAO can be summarized as follows. The study design should be randomized (method stated), double blind and placebo controlled. If possible, monosensitized patients should be selected, but if polysensitized patients are included the causal role of the allergen used for immunotherapy must be clearly ascertained. Standardized vaccines must be used when available and the cumulative dose as well as the maintenance dose (possibly in micrograms of allergen) must be recorded. Rescue medications (as needed) must be prescribed to all patients and recorded on the diary card as well as the total score of symptoms (nasal + ocular + bronchial).

### ***Combined symptom and medication score in immunotherapy trials***

In order to fulfill the need for one and only one final endpoint for trials on allergen-specific immunotherapy, attempts to establish a clinically relevant symptom score, evaluating the potential reduction in allergic symptoms in combination with the potentially reduced need for symptomatic medication, are under investigation. Some examples are mentioned below.

A study by Bufe et al. [3] established a Clinical Index Score (CIS) for a trial investigating the efficacy of sublingual swallow immunotherapy in children with severe grass pollen allergy symptoms.

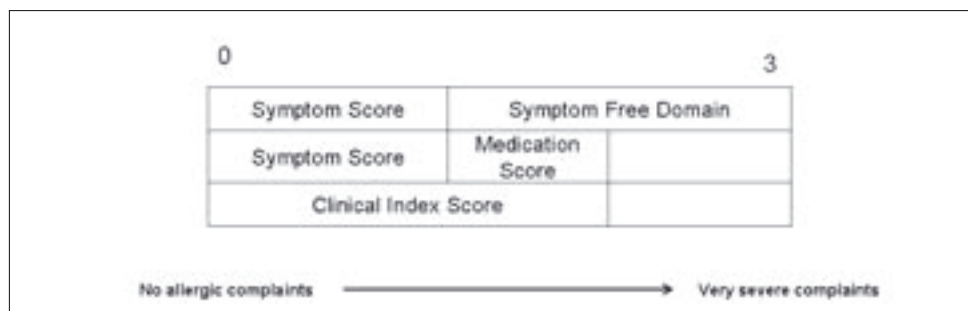


**Fig. 1.** Basic principles of the Clinical Index Score (CIS)

This trial was designed as a double-blind placebo-controlled study on drop-based sublingual immunotherapy and the basic principle of the CIS was defined in the following way (Figure 1).

- The CIS covers a range from 0 to 3 (no to severe)
- If the patient is also using symptomatic medication, the “real” symptoms score would be higher, but can never exceed 3.
- For use of symptomatic medication some value has to be added to the “real” symptom score.
- The amount that must be added is a percentage of the amount that can be added until the maximum of 3 is reached.
- The score provides a tool to calculate the amount that has to be added to the symptom score due to the use of medication (clinical index score).

Weighting of medication and the value to be added to the symptom score was based on an estimate defined by the authors. Local anti-histamines used less than 3 days corresponded to a 10% increase in the “real” symptom score in the eyes and nose, topical corticosteroids 25% in the nose, systemic anti-leukotrienes 10% in the lungs etc.

The conclusions on the CIS score system were that the procedure allowed an analysis of the influence of immunotherapy on the natural course of allergic symptoms and also that CIS turned out to show generally higher scores than the symptoms alone. Furthermore it could be demonstrated that the CIS, which was used as the primary outcome had good responsiveness and an acceptable retest reliability and standard error. The challenges with the Combined Index Score system are to estimate the optimal weighting factor for individual medications and that the system seems to underestimate patients with high-level symptoms.

In another randomized, double-blind, placebo-controlled clinical trial the efficacy of sublingual tablet-based immunotherapy for allergic rhinoconjunctivitis was evaluated using different assessments of combined symptom and medication scores. Five different outcome measures were evaluated and compared [4].

The “Average Rhinoconjunctivitis Total Symptom Score” and the “Average Rescue Medication Score” were calculated and three different ways of combining symptom and rescue medication score were compared. To assess the different outcome measures regarding their power to discriminate between effective treatment and placebo, the effect size was calculated.

The results of this comparison showed that the use of adjusted combined scores provided borderline greater discriminatory power between active treatment and placebo than the use of unadjusted symptoms and medication scores. The highest effect size among the different outcome calculations is associated with the sum score (with equal weights) proposed by WAO [2]. On the scale used, relative low scores between 0.0 and 0.7 were more prevalent for active treatment and higher scores between 0.7 and 1.8 more prevalent for placebo patients. The purpose of this study was to compare different methods for evaluation of data and the outcome should be regarded as explorative more than conclusive.

In a recent clinical development program of a multi-center, double-blind, randomized, placebo-controlled study of the efficacy and safety of a grass allergen tablet, the combined score of symptoms and medication were discussed. In this study, the calculation of a functional relationship between the two variables was tested. In other terms this means that every individual symptom at every different level of severity is changed differently by all different rescue medications at different dosages [5].

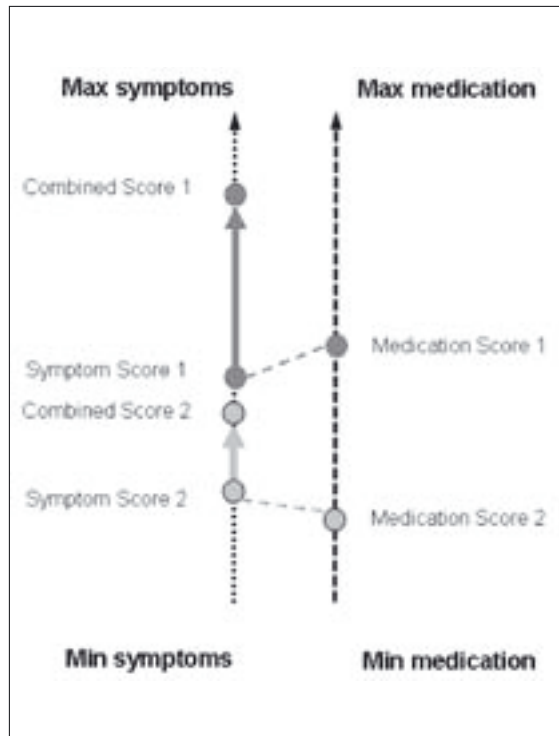


**Fig. 2.** *Functional relationship of combined scores*

A functional system calculates the “true” symptom score corrected for influence of intake of rescue medication. It requires a functional relationship between the use of rescue medication and reduction in symptoms and there is a need for a “symptom-medication matrix” explaining how every symptom at every level is changed by all rescue medications.

The principle and formula for this approach is illustrated in Figure 2 where two hypothetical examples of a functional adjustment are indicated. The formula for the functional estimate is then

$$C = (S/S_{\text{Max}})/(1 - M/M_{\text{Max}})$$

where S = Symptom Score, M = Medication Score, C = Combined Score

In this study, using the described methodology, the combined score, adjusting the symptom score for intake of symptomatic medication, correlates with the pollen exposure and includes the interrelation of the symptom score and the medication score.

## Conclusion

The recently approved EMEA “Guideline on the clinical development of products for specific immunotherapy for the treatment of allergic diseases” has described the issue that the use of rescue medication has an impact on symptom severity [6]. Therefore, the primary endpoint may reflect both, symptom severity as well as the intake of rescue medication. Different approaches to combine symptom score and intake of rescue medication are possible, but in any case the method used to combine both scores has to be pre-specified and justified. So far, no validated system for balancing symptom and medication scores exist. Initiatives for establishing such a balanced and validated scoring system would be helpful for future assessments and should be encouraged and supported by the interested parties.

This paper briefly describes some examples of such initiatives and several projects are ongoing investigating different approaches to finally define the most important single endpoint for allergen immunotherapy trials – an endpoint which measures the reduction in symptoms including the impact of reduced need for daily symptomatic drugs.

## References

1. Note for guidance on statistical principles for clinical trials (CPMP/ICH/363/96), September 1998. <http://www.emea.europa.eu/pdfs/human/ich/036396en.pdf>
2. 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.
3. Clark J, Schall R. Assessment of combined symptom and medication scores for rhinoconjunctivitis immunotherapy clinical trials. *Allergy* 2007; 62: 1023–1028.
4. Bufe A, Ziegler-Kirbach E, Stoeckmann E, Heidemann P, Gehlhar K, Holland-Letz T et al. Efficacies of sublingual swallow immunotherapy in children with severe grass pollen allergic symptoms: a double-blind placebo-controlled study. *Allergy* 2004; 59: 498–504.

5. Jens Strodl Andersen, ALK-Abello 2008, personal communication.
6. Guideline on the clinical development of products for specific Immunotherapy for the treatment of allergic diseases (CHMP/EWP/18504/2006), November 2008. <http://www.emea.europa.eu/pdfs/human/ewp/1850406en.pdf>

## **Discussion**

**Van Wijk:** In your introduction, you talked about “quality of life” as a possible primary endpoint. Perhaps it should not be the primary endpoint, but one of the advantages is that they have a very good estimation of what is clinically relevant and what is not. Referring to those combined symptom scores, you are missing that a little bit – or what is your opinion about that?

**Jacobsen:** The point is that I had to decide here to focus on what my opinion is. You are absolutely right, the advantage of using a rhinoconjunctivitis-specified quality of life measure, as it has been used in many of these studies is that it actually is independent of some of these parameters, like the medication and so on. That is not up to me to decide, but I do not think that just using that as a primary endpoint would be sufficient for a potential confirmatory trial going into a regulatory situation, at least not at this point of time. It could come, it is much more objective as other parameters and it would be of course a good idea to do that. We all know it is there, it is available and it is useful now.

**Durham:** Just along the same lines as our chairperson’s question: It is absolutely right that in practice we are moving away from symptom and rescue medication scores. But of course we are tied to them, not least by the regulatory authorities, as the key primary endpoints. I think we have to be satisfied with symptom scores and rescue medication scores until we have something better. I would also say the global evaluation is quite helpful in the sense that that has a retrospective component and that is very useful. My question, however, is: were you able to combine the symptom and rescue medication scores? The problem that I have come across, particularly from the people who pay the bills, you get very low scores in these studies. So for example on a scale of 0–18 you are getting a rescue medication score of four and a symptom score of 3. And so people say: “Who cares?”. You are not targeting the relevant population because their scores are so low. And of course this is not true. The problem is that symptom and rescue medication scores are always low in immunotherapy trials. It is generic because it is a prophylactic treatment. In other words you got the period in the beginning of the season when the pollen counts are very low and they are very low at the end of the season. They also very intermediate during the season. What we have is an average score. Even in one of Roy Gerth van Wijk’s studies where he selects patients on the basis of a score of greater than 5, I think they had a score of 18. When they were actually monitored during the season, they came up with a score of 3. And this is the problem that we have, I think we do not have to change the studies; we just have to re-educate people that low scores do not necessarily mean mild disease. And I think it is because of all the different things that we see: it is low pollen counts, variable pollen counts, end of season, clinical trial effects, placebo effect, effects of effective anti-allergic medication, producing the scores. It is a perception problem out there that the patients that we are treating have actually

a very mild disease on the basis of their mean symptom scores, whereas in fact they do not. And that is the problem and I think that is why we should be shifting towards more global measures such as rhinitis quality of life.

**Jacobsen:** It is always very convenient when people answer their own questions. Thank you for that, Stephen! Maybe I can just add a little to it and say using a visual analog score could actually be integrated in your symptom and medication score. In the formula you could actually add that to it as it was actually done in the Bufe study. That was a part that I jumped over and did not show you. The other part is that, honestly, I do not think anybody really knows what an average rhinoconjunctivitis patient would score on the scales that I showed you here. Out there in normal life you are interfering with immunotherapy and other things. I think you need such a study because claiming that your patients' diseases are mild because they have a score of 4 on a scale of 18 means nothing because it does not tell you anything about severity. It could be very severe, we just do not know. The scale is arbitrary; it is adding together many different symptoms, and very few patients actually have all of them at the same time, this really is the worst case. Maybe somebody can tell me if anybody ever had a score of 18, I do not think so.

**Pauli:** I have two questions. First, do you think that when you evaluated the total symptom scores of rhinitis and conjunctivitis is it necessary that both scores go the same way? Second, what is your feeling about classical provocation tests like nasal bronchial challenges and the new chamber exposition method?

**Jacobsen:** I think that we are going to hear a little later about surrogate parameters and variables in more detail here and I actually took it out of my talk. Briefly, bronchial provocation or the provocation test can never be a primary variable for a confirmative trial. But it is definitely confirmative to a symptom and medication score. And sometimes I think that especially a conjunctival test is a very sensitive test that will give you an explorative indication even in smaller patient populations. Whether symptoms in the nose and in the eyes always go together, I can only say from a statistical point of view. I have never seen a statistical test in a group of patients showing the opposite. But I do not think I ever looked at the same patient at the same time. So I cannot answer that question.