

Arbeiten aus dem Paul-Ehrlich-Institut
11th International Paul-Ehrlich-Seminar, Bad Homburg 2005
Verlag Chmielorz

Clinical Trials with Allergen Products – Statistical Considerations¹

Peter Volkers

Introduction

Statistical considerations play an important role in the approval of any treatment. This is reflected in several internationally harmonized guidelines dealing with statistical considerations in clinical trials, starting with the ICH E6 guideline [1]. The ICH E6 guideline postulates the need for statistical planning and reporting in any clinical trial. General statistical principles for clinical trials are provided in the ICH E9 guideline [2], while the ICH E10 guideline [3] deals with the choice of controls and design aspects. Specific methodological aspects are discussed in more detail in various guidelines issued by the CHMP or other regulatory agencies. For clinical trials with allergen products among others the following CHMP-guidelines should be mentioned: guideline on missing data [4], guideline on multiplicity [5] and guideline on baseline covariates [6]. However, all these guidelines have in common that they provide general principles. The guidelines are not intended to be used as ‘cook-books’.

When it comes to clinical trials for specific treatments (e.g. immunotherapy with allergens) or specific indications (like allergic rhino-conjunctivitis) some principles mentioned in these guidelines apply directly while others need a more in-depth consideration in order to translate the general principles to the specific trial situation. This is especially important in case of confirmatory clinical trials, intended to support a marketing application authorisation. However, even in earlier phases of a drug development a careful assessment of statistical aspects is required, as the validity of results of early studies impact later stages of drug development.

Statistical requirements that apply to all confirmatory clinical trials are: the intended statistical analysis has to reflect the (clinical) objectives of a study proper-

¹ The view expressed in this article is the personal opinion of the author and is not the official opinion of the Paul-Ehrlich-Institut

ly and the statistical analysis has to be pre-planned [2]. However, translating these general principles to clinical trials with allergen products entails some consequences on study design and analysis planning.

In order to discuss some of the problems with statistical considerations in clinical trials with allergen products, confirmatory efficacy trials for the treatment of seasonal allergic rhino-conjunctivitis (SAR) will be used as example.

Starting with a discussion of statistical aspects related to the definition and documentation of the study endpoint, additional aspects like study design, sample size and analysis planning will be briefly discussed. Hopefully it will become apparent that clinical and statistical considerations have to complement each other in order to maximise the chance for a positive study outcome.

Endpoint considerations

According ICH E9 the primary endpoint should be the variable providing the most clinically relevant and convincing evidence directly related to a clinical trial.

Furthermore,

- the primary endpoint should reflect accepted norms and standards
- the use of reliable and validated variables (as primary endpoint) is recommended
- the primary endpoint should allow for a valid and reliable measure of a clinically relevant treatment benefit
- the primary endpoint should generally be the one used for sample size estimation.

All these requirements are generally acceptable. However, applying them to a clinical trial with an allergen product is not as simple as it might seem. In most clinical trials with allergen products, the primary objective is to improve the symptoms of the allergic disease. Usually these symptoms are measured by means of a symptom score assessing the severity of different symptoms of e.g. rhino-conjunctivitis on a daily basis over a certain time period. The use of patient self-rating scoring systems is standard.

However, bearing in mind the requirements on primary endpoints in a clinical trial, there are different aspects with this endpoint that have to be assessed critically:

- In most trials with allergen products the use of rescue medication is allowed. However, the use of rescue medication has an impact on the severity of symptoms and thus the symptom score.
- Up to now, there is no generally accepted symptom score and most of the symptom scores used in clinical trials are not validated.
- As a consequence, for most scoring systems there is no generally accepted definition of a 'clinically relevant benefit'.

The use of rescue medication indicates insufficient control of symptoms. Thus, a symptom score not accounting for any use of rescue medication is not a direct meas-

ure of a treatment effect, but a measure of the combined effect of treatment and rescue medication used. Therefore, in order to allow for an unbiased assessment of the treatment effect, the use of rescue medication has to be considered in the statistical analysis. From a statistical perspective one might think about different approaches (it has to be mentioned that other approaches are also possible):

- Both symptom severity score as well as rescue medication use are assessed as a co-primary endpoint. A possible statistical testing strategy would test the new allergen product for superiority vs. the comparator with respect to the symptom severity score and in parallel would provide evidence that the use of rescue medicine is not relevantly increased for the new allergen product relative to the comparator. However, such an approach imposes problems, as it requires a valid documentation of the daily medication consumption as well as a valid quantitative measure to compare medication consumption (probably a scoring of different types and/or dosages of rescue medication) and a comprehensible definition how to quantify, that the use of rescue medication is not relevantly increased. Furthermore, the use of more than one endpoint requires special statistical considerations to avoid an inflation of the type I error in a clinical trial [5]. The use of two co-primary endpoints might have also implications on the sample size required for a clinical study.
- Another possibility to account for the intake of rescue medication would be to add some kind of penalty to the symptom score every time a patient makes use of rescue medication. This approach avoids some of the (statistical) problems associated with the use of co-primary endpoints. But still, the problem remains how to quantify the use of probably different types and dosages of rescue medication, how to define a 'correct' penalty for the use of rescue medication and how to combine both (symptom and medication score) to a valid combined score. Such an approach (however without explicit solution) is mentioned in the guideline for treatment of allergic rhino-conjunctivitis [7] (although this guideline does not deal with specific immunotherapy).

Thus, from a statistical point of view there is a need to investigate more in the definition of an appropriate endpoint that properly account for the use of rescue medication, the quantification of different types or dosages of rescue medication being one of the major challenges for defining such endpoint.

However, regardless whatever solution will be used to account for rescue medication, the symptom score itself poses some challenges for the statistical planning and analysis of clinical trials with allergen products. Recently there is a whole 'universe' of different symptom scores used in clinical trials. Different scores are to be found not only in different development programs, but even within the same development program. This has obvious implications:

- Often treatment effects are not comparable across studies with comparable targets as different scales are used
- The use of different scales complicates sample size estimation for new clinical trials.

From a statistical perspective, the question whether symptom scores, used in clinical trials with allergen products, allow for a valid and reliable measure of clinically relevant treatment benefit is still open and further clarification on how to define such benefit is highly desirable.

Design considerations

The choice of a suitable study design is, besides the choice of the primary endpoint, a major statistical issue when planning a clinical trial. The variability of allergen exposure, the spontaneous variability of the disease treated as well as the subjective nature of any symptom score have to be carefully assessed when designing a clinical trial for allergen products.

The major aim of any statistical planning in a clinical trial is to allow for an unbiased assessment of treatment effects. Due to the different sources of variability in clinical trials with allergen products, only the use of a comparator (either placebo or another active substance) allows for an assessment whether observed effects are due to the natural course of the disease or whether these effects are due to other factors like the environment, or whether they might be attributable to the treatment under investigation. However, performing a comparative trial is not sufficient to avoid bias. There are various sources of bias that have also to be taken into consideration when planning a clinical trial. Among others the following types of bias have to be considered:

- *Selection bias*: Knowledge of the treatment of the next patient might influence the decision to enter a patient into a clinical trial, possibly introducing an element of imbalance between treatment groups.
- *Performance (treatment) bias*: Knowledge of treatment allocation might (intentionally or unintentionally) lead to differences in patient care or the patients' behaviour.
- *Assessment bias*: Knowledge of treatment in a comparative trial might have an impact on the assessment of efficacy and/or safety parameters. This is especially true in most trials with allergen products as in these trials (self-) rating scales are usually the primary endpoint.

Randomisation is the method of choice to deal with selection bias. However, the randomisation procedure applied has to ensure that the investigator is not aware of any treatment of future patients to be included in the study (thus providing investigators with an (open) randomisation list is not appropriate). Proper randomisation may be achieved e.g. by centralised (telephone) randomisation. If properly performed randomisation should ensure that the distribution of known and unknown prognostic factors is similar between the treatment groups in a clinical trial. In addition the study protocol should standardise the treatment conditions as much as possible.

Performance (treatment) bias can only be minimised by blinding the participants of a clinical trial. This includes also the Sponsor's staff in order to avoid any possible dissemination of treatment allocation.

Blinding is also the most suitable measure to minimise assessment bias. This is especially true in case of 'subjective' endpoints (like self-scoring systems). Only in

case of objective 'hard' endpoints the use of so-called PROBE-designs (open trials with blinded endpoint assessment) might help to minimise an assessment bias. However, with a PROBE design the problem of a possible performance bias still remains. For clinical trials with allergen products and subjective assessments it is questionable whether results from a PROBE-design are unbiased.

To summarise: The optimal way to deal with bias in a clinical trial with an allergen product is to perform a double blind, randomised comparative study.

Due to

- the variability of the disease,
- the variability of environmental factors (e.g. pollen count) that may influence study results and
- the problems with defining a clinically relevant treatment benefit based on symptom (or symptom-medication) scores,

clinical trials with allergen products should be planned as superiority trials. Two-armed non-inferiority trials (comparing a test product to an active comparator with respect to efficacy) are recently not acceptable due to the lack of assay sensitivity, related to the factors mentioned above. If for any reason, the equivalence or non-inferiority of a new allergen product, compared to an already authorised product, is the aim of a clinical development program, the pivotal studies in this program should be performed as three arm trials, including a placebo arm, in order to allow for an assessment of assay sensitivity.

Sample size considerations

Sample size consideration is one of the most important aspects when planning a clinical trial. The sample size of a clinical trial should provide sufficient likelihood to allow for a conclusive study result. A sample size not allowing for a valid study conclusion may lead to a waste of money. Furthermore, in an underpowered clinical trial subjects participating in this trial are unnecessarily exposed to risks.

Therefore, sample size considerations are mandatory for all clinical trials (ICH E6). In order to allow for a valid sample size calculation one has to make assumptions: Besides technical / statistical issues like the choice of the statistical model used for analysis, the Type I error (usually 0.05 2-sided or 0.025 1-sided) and Type II error (20% or lower corresponding to a power of at least 80%) respectively, assumptions with regard to the primary endpoint have to be made (according to ICH E9, the primary parameter should be used for sample size calculation). Assumptions on the variability of the primary endpoint as well as on the treatment effect (e.g. difference in means) to be detected are necessary to perform such a calculation. Both, the assumptions on variability as well as on the treatment effect to be detected should be carefully considered, as they have the largest impact on a proper sample size. As long as there is some experience with the same endpoint from earlier trials the data from such trials can (and should) be used when making any assumption on the variability of the primary endpoint. However, with respect to the treatment effect to be detected, the situation is a little bit different. According to ICH E9, "...the treatment

difference to be detected (and thus used for sample size calculation) may be based on a judgement concerning the minimal effect which has clinical relevance....". Thus, the problem remains how to define a minimal clinically relevant difference for symptom scores used in clinical trials for allergen products.

A sample size that provides a proper chance for a conclusive study outcome according to pre-specified criteria is important. As the information necessary for sample size calculation is often vague, sample size adjustment is a feature applied in many clinical trials. However, even if such adjustment seems possible (although it might not be desirable) one should carefully reconsider the need and implications of such procedures. Some of the methods for sample size adjustment have implications on the study design as well as the study analysis, as they require unblinded interim analyses. But, as any interim analysis might endanger the integrity of a clinical trial, methods for blinded sample size adjustment (e.g. Kieser and Friede [8]) are preferable (at least from a regulatory point of view). However, there is a specific problem for sample size adjustment in clinical trials with allergen products. If these trials deal with a disease (e.g. SAR) that has a seasonal manifestation, any sample size adjustment based on study data usually implies a prolongation of the study for at least one additional season.

Analysis considerations

From a statistical perspective, bias is one of the major (if not the major) concerns with clinical trials. Bias can occur not only during study conduct, but also when analysing study data (such kind of bias is often called analysis bias):

- Different methods might be used to analyse a data set. For example, one might use either a two-sample t-test or a Wilcoxon-test to compare two groups. Analysis bias occurs if one performs both tests but reports only the results of the test that is most in favour with the desired result.
- Analysis bias may also occur with the choice of the analysis population. The exclusion of e.g. patients from analysis might impact the direction or extent of study results.

In order to minimise a possible analysis bias, the major features of the statistical analysis of a clinical trial should be pre-specified in the study protocol (providing detailed statistical methodology at least for the analysis of the primary endpoint). An extensive statistical analysis plan should be finalised prior to unblinding the study (ICH E9).

It is taken for granted that the statistical hypotheses to be tested are in line with the study objective(s) and appropriate statistical methods are applied. For example, to conclude equivalence or non-inferiority based on a non-significant study result of a test for superiority is inappropriate. Furthermore, a statistical analysis providing only p-values is not acceptable. In order to allow for a clinical judgement, treatment effects (e.g. differences of means, odds ratios etc) including appropriate confidence intervals have to be provided.

While the principles mentioned above are not specific to clinical trials with allergen products, there are some issues that raise difficulties when applied to such trials and that have to be carefully considered.

Analysis population

In superiority trials the intent-to-treat principle should be applied, i.e. all patients randomised should be included into the analysis according to the treatment randomised. This approach is an additional measure to minimise bias. However, in clinical trials with allergen products it might be hard to follow the ITT principle for two reasons. One is the long study duration as well as the time between treatment start and onset of 'environmental challenge' (e.g. for SAR). Both aspects may result in a relatively high number of patients determining a trial prematurely. Provided that the primary endpoint is documented, these patients (no matter whether they were treated according protocol or not) can (and have to) be included into the analysis population for the primary endpoint. However, this requires that even in those patients that terminate a study early, any effort is made to collect the information on the study endpoints (which might be hardly possible in reality).

In e.g. SAR-studies the environmental burden is another aspect that might impact the definition of the analysis population. To allow for disease symptoms a suitable provocation is necessary. Without such environmental burden none or only minimal symptoms occur. This impacts the informative value of any clinical study in this indication. Thus a modification of the ITT population might be meaningful. For instance, one might think about the exclusion of all (not only single) patients from a centre if the pollen count for this centre falls below a predefined threshold, throughout the whole pollen season. However, any criterion for modifying the ITT population has to be transparent, easy to understand, clinically justified and has to be clearly pre-described in the study protocol. Furthermore, one should be very careful when thinking about such a modification of the primary analysis population as this might impact the power of statistical tests planned for analysis, possible claims for indication and the acceptance of the study by regulators. Thus, in case a modification of the ITT population is intended, regulatory advice should be sought for when planning a clinical trial.

Missing data

Missing data are a common problem with all clinical trials. However, compared to clinical trials with other therapies, the problem of missing data is more pronounced in clinical trials with allergen products. This is mainly due to two reasons; the relatively long treatment duration (often starting at a point in time where patients do not experience any disease symptoms) as well as to the endpoint of such studies (documentation of different symptoms in a patient's diary over a longer time period).

Roughly speaking, there are two types of missing values: Missing values where there is no evidence that the absence of data is treatment related and missing values where it cannot be excluded that there is a relation between the absence of data and

study treatment. For missing data where there is no evidence of any relationship between missingness and study treatment there is a number of procedures available to substitute for missing values. Common approaches to deal with missing values are linear interpolation (for data missing ‘in between’) or the last observation carried forward approach (i.e. the extrapolation of the last observed value). However, any substitution of a missing value might bias the study results. Thus in confirmatory studies further sensitivity analyses might be necessary. For details refer to the guideline on missing data [4].

While there are different approaches to deal with missing data that are not treatment related, treatment related missing data pose a very severe problem to any analysis. In SAR trials such data may occur if patients terminate the trial due to undesirable adverse effects during the initial therapy phase (i.e. prior to the start of a pollen season) without documenting their symptoms during the following pollen season. Under these circumstances, a relationship between treatment and the absence of data cannot be excluded. An exclusion of such patients from the analysis (as often done) will bias the study result. The same problem (even more pronounced) occurs in case a patient terminates a study prematurely, due to a lack of efficacy without documenting the primary endpoint. As in clinical trials with allergen products the number of patients that terminate the trial prematurely often is quite high (compared to trials with other therapies) and it cannot be excluded that a number of these premature terminations are treatment related, efforts have to be made to deal with this situation. One possibility (although a more theoretical one) is to collect the primary endpoint from all patients (regardless, whether they are on therapy or not). An alternative approach could be to account for absence of data in the definition and analysis of the primary endpoint. However, such an approach (e.g. according to methods outlined in a different context by Chang *et al.* [9]) needs further work, as it cannot be solved within the framework of ‘standard’ methods often applied in the analysis of clinical trials with allergen products.

Covariables

The inclusion of prognostic factors known to influence the primary endpoint into the statistical analysis model usually increases the efficiency of the statistical procedures. However, it has to be mentioned that only factors that are not treatment-related should be used as covariate in a statistical model. For treatment related covariates their inclusion into a statistical model may result in biased results.

Examples for known prognostic factors in clinical trials with allergen products are symptom severity at baseline (i.e. measured prior to randomisation) or pollen counts during the pollen season. Accounting for continuous covariables like symptom severity at baseline, is usually done by means of an analysis of covariance approach (ANCOVA, assuming normally distributed data). This method can also be applied to account for pollen count (although it might be impossible to document the pollen burden for each patient). However, a simpler method to account for pollen count during the pollen season is an analysis stratified by the centre (assuming that the pollen burden is similar for patients from the same centre). This method can be

applied in parametric as well as in non-parametric analyses. If such a stratified analysis is planned, the randomisation should already account for this stratification (in order to maximise the efficiency of the procedure).

While the use of covariates might increase the efficiency of a statistical analysis, it has to be mentioned that only prognostic factors for which a clinical rationale exists should be used. Furthermore covariables should be prespecified. More information on methodological aspects with covariables can be found in the guideline on adjustment for baseline covariates [6].

Conclusions

Although statistical considerations are important for the success of a clinical trial, there are still a lot of issues that have to be investigated in more detail in clinical trials with allergen products. This holds in particular for such aspects, which arise from the specific features related to immunotherapy with allergens. Among others the following problems need further consideration:

- Suitable definition of a primary endpoint, allowing for an unbiased estimate of the true treatment effect
- Careful assessment of the statistical characteristics of the primary endpoint (e.g. distribution characteristics)
- Suitable preplanning to account for missing values or their consideration in the primary endpoint
- Inclusion of prognostic factors into the statistical analysis model (e.g. baseline severity, environmental burden).

The problems mentioned above are not solely a question for statisticians, but have to be solved in close collaboration between clinicians and statisticians.

References

1. ICH Topic E6: Guideline on good clinical practice. CPMP/ICH/135/95.
2. ICH Topic E9: Statistical principles for clinical trials. CPMP/ICH/363/96.
3. ICH Topic E10: Choice of control group in clinical trials. CPMP/ICH/364/96.
4. Points to consider on missing data. CPMP/EWP/1776/99.
5. Points to consider on multiplicity issues in clinical trials. CPMP/EWP/908/99.
6. Points to consider on adjustment for baseline covariates. CPMP/EWP/2863/99.
7. Guideline on the clinical development of medicinal products for the treatment of allergic rhinoconjunctivitis. CPMP/EWP/2455/02.
8. Kieser M, Friede T. Re-calculating the sample size in internal pilot study designs with control of the type I error. Stat Med 2000; 19: 901–911.
9. Chang MN, Guess HA, Heyse JF. Reduction of burden of illness: A new efficacy measure for prevention trials. Stat Med 1994; 13: 1807–1814.

Discussion

Malling: Thank you very much. I agree with almost all of your points. I think that it is extremely important that when we plan immunotherapy studies we have to take into account statistical considerations. Because otherwise we will not be able to come to the right conclusions.

Bufe: I have not quite understood why non-inferiority studies cannot be done with this sort of settings. If the number of patients you look at is high enough, if the power is high enough, why can you not do this? Is it too variable?

Volkers: Well, I think it is too variable. The problem really is, and I think you all have experienced this, while in one year a drug works whereas when you repeat the trial in the next year, it does not work. And this is only due to the fact that, e.g., there are no pollens around, there is no environmental burden.

Bufe: Okay, but if the endpoint is something independent of that?

Volkers: If the endpoint is valuable and validated, I think then it is possible. But right now we do not have such an endpoint and therefore it is not yet possible.