

Adequate Clinical Trials of Immunotherapy

KJELL AAS

The Allergy Unit, Pediatric Research Institute, Rikshospitalet, University Hospital, University of Oslo, Oslo, Norway

1. Introduction

The number of papers concerned with allergen specific immunotherapy (hyposensitization) has become quite considerable. Enormous resources are put into investigations on the clinical and immunological effects of various hyposensitization or pharmacotherapeutic regimens with the use of different material. With continuation of the extensive practice of hyposensitization, this kind of research is highly needed and appropriate, provided that it is adequately carried out and properly reported. The frustrating fact is, however, that very few – if any – of the published trials appear completely adequate to the critical observer. Many of the manuscripts on hyposensitization are rejected by reviewers and editors of journals of high standard. Some of these papers remain unpublished, but others may reappear with little alteration in other journals with less critical referees and editors. Even this journal has inadvertently printed paper(s) about clinical trials which ought not have been published in the form submitted.

The critical reader is left with feelings of irritation and frustration (1, 27, 29).

The impression given by generally accepted literature is that hyposensitization is effective under optimal conditions. Thus, it is not hyposensitization as a therapeutic principle that should be put on trial, but the quality of the extracts employed – and in some instances also the qualifications of some of the physicians working in this field.

The present review aims at giving guide-lines for adequate trials of immunotherapy. Similar principles apply to the clinical trials of drugs. The review is meant not only for those planning investigations of this kind, but may hopefully serve as a guide for those who want a yardstick for a constructive evaluation of available literature. However, a review of this kind must not be considered an attempt to dictate rigid rules. On the contrary, it represents only a humble attempt to communicate ideas and the experiences gained during years of basic and clinical trials with many failures and some successes – and reflections made after having read many papers on inadequate trials.

2. Selection of patients and controls

Accepted indications and no contra-indications for hyposensitization must be present for all patients and controls included in the study (20).

The scope of hyposensitization is limited to reagin-mediated diseases such as allergic rhinitis and allergic bronchial asthma, presenting major problems to the patient. Hyposensitization should only be considered when it is shown beyond doubt that the particular allergen(s) play(s) a major role for the symptoms and severity of the disease and the allergenic substances in question can not be sufficiently eliminated in spite of efforts to accomplish this.

The criteria as to the importance of allergy for the disease in question should be clearly defined. It is not sufficient to demonstrate the

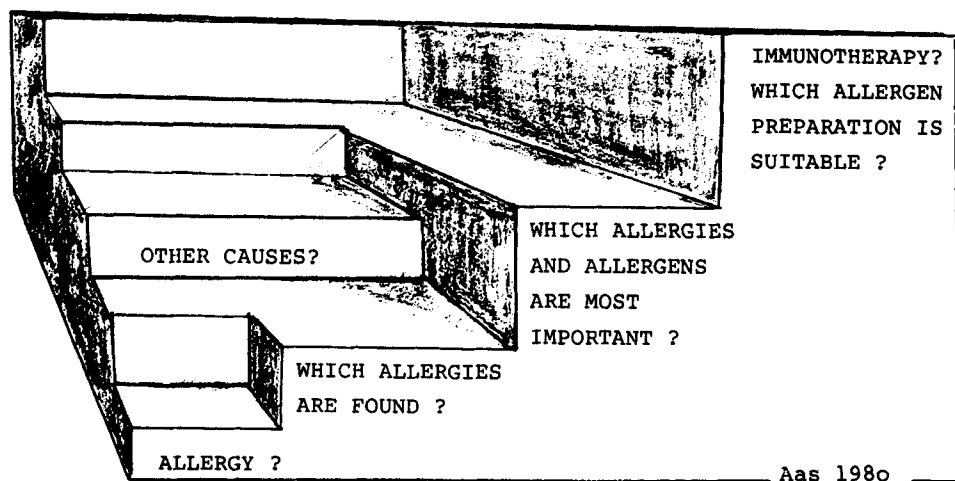


Fig. 1. The diagnostic fundament for the therapeutic approach. Both non-allergic and allergic causes should be considered. Many diagnostic steps are needed before hyposensitization is commenced.

mere presence of the particular allergy by means of skin testing and/or RAST only. The diagnostic screening must document that the allergens in question are crucial to the disease and the reaction in the affected tissues (Fig. 1). A large number of patients with seasonal allergic rhinitis and bronchial asthma, for example, may show positive skin test and RAST reactions to both pollen and moulds. Pollen allergy may play a major role for the rhinitis, and mould allergy may be the major cause of bronchial obstruction in some of them. Hyposensitization with pollen extracts alone would not be of much benefit as regards asthma for this kind of patient.

Parallelism between atmospheric pollen counts and asthmatic symptom scores may not be sufficient diagnostic evidence for pollen allergy being the major cause. Mould spore counts may many times parallel the counts for pollen during the season in question.

Children with mild or very moderate asthma are not well suited for trials of this kind. They can improve so much in a few years, simply due to the natural course of mild asthma, that the demands for adequate numbers of non-treated controls and for objective measures may present insurmountable obstacles to obtaining statistically significant results. Hence, more severe

cases with a less favourable prognosis when untreated are to be preferred. The criteria as to this should be defined, and are based on simple clinical observations (5).

Control individuals should be matched with the treated patients with respect to the parameters that are most important in the investigation. Most papers on hyposensitization fail to satisfy this demand. Matching with respect to age and sex is frequently the only parameter reported, but may not be appropriate. It is more important that the patient and control individuals are comparable with respect to type, severity and prognosis of the disease, role of the allergen in question and degree of both specific and non-specific reactivities.

Emphasis should be given to parameters such as pretreatment symptom and drug consumption scores related to the degree of allergen exposure, nasal or bronchial reaction thresholds to the allergens and other agents, and so on. If two different allergen preparations are to be compared, matched patients and control individuals should show the same degree of allergic response to challenges with both preparations in an optimal trial, or at least to the particular allergen preparation used for treatment and evaluation (by challenge tests).

The number of patients and controls in-

cluded in the study should be decided from both practical and statistical viewpoints (see section 11).

If immunological parameters are emphasized (IgE, blocking antibodies, histamine release, suppressor T-cells) a non-treated control group should be included that initially is comparable for these parameters. Matching may become difficult if many parameters are investigated at the same time. If this is the case matching should be attempted following a list graded as to the importance of each component, and this should preferably be defined and discussed in the presentation.

3. *Informed consent*

According to the Helsinki treatise on ethics in medicine (20, 26) all individuals should receive adequate information about the trial including anticipated effects and side-effects. They should give their written, informed consent to participate. For children it may be necessary to have informed consent from both parents, but also the child has the right to be informed and to influence the decision on his/her own premises (14, 19). The information given must include information about the necessary assessment procedures during the whole trial. It is quite demanding for the patient and control individuals to participate in an adequate, prolonged trial of hyposensitization, and great pains must be taken to inform and motivate the patient. If the subject is in a dependent relationship to the clinical investigator, informed consent should be obtained by a well-informed physician who is not engaged in the investigation (20).

If each individual is not adequately informed of his statistical risk of receiving ineffective treatment, a double blind placebo-controlled trial must clearly be deemed unethical. When adequate information has been given, it may be hard to raise enough motivation in the individuals to go through a long period involving all the necessary assessment procedures and a therapeutic or trial programme. Lack or loss of patient motivation may make otherwise perfect trial protocols unfeasible.

With the above reservations in mind and – in particular – with consideration of the ethical

barriers, it may appear justified to investigate the effects of hyposensitization in open trials. Mandatory provisions for the open trials should be that the most reliable objective recordings are used to assess the effect, and that well-matched non-hyposensitized individuals are used as controls. If neither provision can be satisfied, there is no sound basis for a clinical trial.

4. *Active material*

Hyposensitization should be given only with preparations from allergen sources crucial to the disease in question.

In seasonal allergic rhinitis this may not present so many problems since pollen from wind-pollinated trees, grasses or weed are the major causes in most instances, and in several patients only one type of pollen may be of importance during the particular season. For such patients the question is often limited to that of presence or absence of cross reactivities between the atmospheric pollens found during the troublesome season. This could in itself present adequate grounds for comparing injection treatment with extracts from one grass pollen and extracts from several pollens. At present the weight of evidence indicates that hyposensitization with one grass species (timothy) provides specific relief for patients allergic to several grass species (15, 23). This indicates a high degree of cross reactivity between the most common and important grass pollens, whereas weeds and wind-pollinated trees show much less cross reactivity (16).

In allergic asthma more problems may be raised by the presence of a bronchial allergy to several different allergen sources, such as various moulds, house dust mites, and animal epithelium in addition to pollen. Most investigators have preferred to select only one allergen source for trials and included patients stated to be predominantly allergic to that particular allergen source. Documentation as to this is often inadequate, however, since the allergen extracts and assay systems used may have been inappropriate for making this distinction (see section 2). Such documentation is essential for all investigators using symptom scores and drug

consumption scores for effect evaluation, but is less important when objective measures are (and can be) relied on.

If there should be reasons to include several types of allergens in the trial, difficulties as to assessment of therapeutic effects may be partly overcome if objective measures can be relied on to replace symptom scores.

One essential factor for a study of that kind is that the allergenic qualities and total allergenic activity of the extracts used remain unchanged between the initial and later challenges. One must consider the limited shelf-life of the preparations and extracts used. Batch-to-batch variations as to total allergenic activity and allergen composition may have a tremendous impact on the results obtained. Even two lots of a pollen extract from the same company have shown a 100-fold difference in major allergen content (6, 15).

If the extract used in the final challenge for evaluating the therapeutic effect is much stronger than the extract used in the initial tests, a conclusion of no or little effect on tissue threshold is likely. If the extract used at the final challenge is much weaker, the tolerance thresholds would seem to be much higher, and this might be attributed to the effect of therapy. Failure to document that the extract's biological activity is approximately the same at the time of challenges after 1 to 2 or 3 years is a main objection to most papers on hyposensitization. From personal experience this reviewer knows that editors have omitted information about this in order to save page space, but such an act of manuscript mutilation is inexcusable.

A few years ago the only available control measures as to the allergenic activity of the extract were skin test titration and – for a few preparations – quantitation by means of immuno-precipitation of defined antigens in the extract. Nowadays, allergen extract standardization can be accomplished more precisely (2, 4). Furthermore, the availability of purified, well-specified and standardized preparations provided in a freeze-dried form may reduce the problem of variability of allergenic activity to a minimum (4, 23). The investigator can him/herself check the stability of the allergenic activity

of the preparations used by combinations of skin test titrations and – if available – RAST inhibition.

The allergen preparation used in hyposensitization should be proven to be active in producing the allergic symptoms in the tissues in question by means of nasal or bronchial challenges with appropriate doses and concentrations. If not, the preparation cannot be considered to be much more than a rather complex placebo preparation. This calls for use of the same preparation for diagnosis and therapy, or at least that the two have identical composition with respect to the allergenic components.

Other characteristics of the allergen extracts used should also be taken into consideration for trials on hyposensitization (6). Allergen extracts used in treatment should be considered in the same way as drugs. As such, the extract should contain defined concentrations of the active components and as little as possible of other contaminating components. For drugs the removal of impurities besides the active components and necessary carrier substances is indispensable.

5. *Use of placebo*

Patients receiving placebo in place of allergen injections frequently show some improvement. This is not necessarily due to beneficial psychosomatic effects. Changes in the disease over a period of time may reflect the natural course of the disease, the more or less controlled fluctuations in allergen exposure (as for instance differences in pollination and sporulation from season to season), or the influence of more or less undefined trigger-stimuli of different kinds.

Injection treatment, be it extract or placebo, is only adjuvant to several other therapeutic measures. The injections may indeed be indifferent components in an otherwise effective multipotent treatment.

Treatment also involves various psychosomatic mechanisms including the physician's own influence on the patients' attitude to the disease. Parameters for evaluation of the treatment such as days missed at school or work,

amount of medication used, and diary score for symptoms, may be as much indicators of changes in the patients' attitude to the disease and the effect of simultaneous therapy, as they are benefits from injection treatment.

In trials depending mostly on a subjective assessment of treatment, it may be mandatory to include placebo. When the assessment relies mainly on objective measures, however, there seems to be less need for a placebo group. In a trial designed as a double-blind controlled study, fewer individuals are necessary to achieve results which are statistically significant for effective treatments.

Many seem to think that all types of therapeutic trials have to follow double-blind designs with the use of a placebo-treated control group in order to make the results reliable and the trial acceptable. This has recently been challenged in authoritative studies in highly rated scientific journals (21, 36). Use of placebo may raise both practical and ethical problems (26). The scientific documentation in favour of hyposensitization with good allergen extracts by far outweighs that against, e.g. with pollen, house dust mite, cat dander and insect venom (1, 3, 25, 29, 30). With such allergens it may appear unethical to keep a group of patients on a prolonged injection regimen with placebo instead of appropriate allergen preparations. As stated above it may become impossible under these circumstances to achieve cooperation and motivation in a sufficiently large number of correctly informed patients. This causes a conflict of interest between what is best for the patient and what is best for research.

Use of placebo may have its proponents, particularly when it is a question of immunotherapy trials with extracts from moulds and some animal dander allergens. There exists no scientific documentation of the efficacy of hyposensitization with mould allergens. It may be argued, however, that the effects of hyposensitization must be principally the same for all allergens and that lack of therapeutic effect with, e.g. mould extracts, is simply explained by the lack of acceptable extracts for the purpose (6). This view opposes the use of placebo in all kinds of hyposensitization since all patients have the

right to effective therapy even in therapeutic trials (20, 26). One may have to compromise in this dilemma.

If placebo is used in a double-blind study, it must be so similar to the genuine therapeutic agent that no difference can be identified with respect to appearance or immediate effects (colour, pain, taste, smell). This is almost impossible to achieve during hyposensitization where typical local reactions occur shortly after the injection of allergen extract when the highest tolerated dose is approached. In order to mimic this, some investigators have added histamine to the placebo solution (38). It would also be necessary to add some injectable colour substance to the placebo to make it resemble the highest concentrations of some extracts. Using such contaminated placebo may not be recommendable, and raises further ethical problems. The same problems are met with when two or more allergen extracts are to be compared in hyposensitization. We found this an insurmountable obstacle in studies comparing the therapeutical effects of crude extracts from timothy, and from a mixture of grass pollen, with those of a highly refined freeze-dried timothy pollen preparation (23). To make the purified preparations similar in appearance to the crude extracts, the former would have to be contaminated with colouring substances and irritants!

In clinical practice, however, it is not a question of whether hyposensitization should compete in efficacy with no-treatment (placebo). The question is whether hyposensitization has a rational place instead of, or in addition to, drug therapy. The pros and cons of hyposensitization are balanced against those of drug therapy (Fig. 2). In other instances it may be a comparison between various allergen preparations or different regimens of administration. In this situation it may be appropriate to use a double-blind system with placebo for either allergen extract or drug in a combined pharmacotherapeutic/immunotherapeutic trial.

6. *Diagnostic procedures and assessment of therapeutic effect*

The aim with hyposensitization is to reduce or

	IMMUNOTHERAPY (Hypo)	PHARMACOTHERAPY (Drug)
Cause oriented	↑ +	-
Allergen specificity	↑ +	- ↑
IgE synthesis	↑ Reduced	Increased
Effect duration	↑ Prolonged	Short
Effect localization*	↑ General	(Local)
Administration*	↑ Demanding (Physician)	Easy ? (Self)
Administration* (frequency)	↑ Weekly - Monthly	3-4-times daily
Time needed for evaluation	6 months - 1 year	Days - Few weeks ↑
Mechanisms of action	?	+ - ?
Side effects, acute*	- +	+ -
Side effects in the long run	-	- ?
Reliability	60 - 85 %	60 - 85 %
Material used	↑ Undefined/ Defined	Defined ↑
Monitoring methods	-	- (+)
Total cost*	High	Moderate / ↑ High

ADVANTAGE ↑

INDIVIDUALLY DECISIVE *

Fig. 2. Advantages and disadvantages of hyposensitization and drug therapy. In each individual case the therapeutic approach is determined after a balanced consideration of pros and cons as indicated by the above "Duellogram".

ablate the allergen-specific immunological trigger stimuli causing allergic disease. For this purpose, methods are needed which are appropriate for measurement of the type, degree and impact of the immune mechanisms in question. Such methods would be most suitable for the diagnostic selection of the type of treatment to be given and for monitoring effects of immunotherapy (see section 12).

There is, however, no immunological method at present and indeed only a few clinical methods can be considered decisive in the selective process before hyposensitization trials. In any case, these are the same methods as those which should be employed in clinical practice prior to hyposensitization. For trials of hyposensitization, however, the methods have to be optimized, made more precise and reproducible, and have to be designed for quantitation as far as possible (4).

In principle the initial diagnostic stage leading to hyposensitization encompasses evidence that the tissue reactivity to the allergens in the therapeutic extract is of decisive importance for the disease in question. To provide this evidence, several diagnostic steps have to be taken, each subsequent step more demanding than the previous (Fig. 1).

When it is only a question of yes or no as to sensitization, there are few demands as regards test procedures and material quality. This is not so in clinical research in which the demands may be rather difficult to satisfy. The investigator must show that the allergens are of decisive importance for the disease and the most crucial ones compared with other relevant allergens and non-allergic causative factors. This means that the tests used must be performed with the highest possible degree of precision, and the material must be of high quality. All

tests in this type of work depend on material that has the same activity at all stages of the test.

7. *Is the extract suitable for the individual patient?*

In hyposensitization it is a matter of course to use allergen extract(s) that represent the most important allergenic sources causing disease in the individual. Inappropriate injection therapy may induce clinical sensitivity to allergens to which the patient is not already sensitive (24, 39). Too impure and complex extracts may be less effective than refined ones, and may even be harmful to the patient (6). This may not be any problem when it is a question of allergy to pollen. Most patients reacting to a given type of pollen are allergic to the same major and intermediate allergens present in a high quality extract of that pollen. Such extracts may therefore be suitable for more than 90% of the patients. It should, however, be confirmed that the patient does react specifically to the extract (preparation) intended for use in therapy. This procedure is simplified by consequent diagnostic use of the preparations most frequently employed in immunotherapy.

When it is a question of allergy to other sources such as animal dander and/or moulds, the pattern of reactivity to different components of the extracts may vary a lot between individuals, and a given commercial extract may be far from suitable for a large proportion of the population allergic to a particular mould. This indicates the need for a final diagnostic step before hyposensitization to find out if a given extract for treatment is suitable for the individual patient in question. It is not likely that the minor allergens have any influence on the effect of hyposensitization, but available information indicates that major allergens, and probably also some of the intermediate allergens are important (6). The content of major and intermediate allergens in the treatment extract should henceforth be shown to bind IgE-antibodies present in the serum of the patient to be treated. This would demand the inclusion of a crossed radioimmuno-electrophoresis study in the final diagnosis before initiating hyposensitization with the extract.

8. *Challenge tests*

Challenge tests with the relevant allergens are the most direct and only available objective means of assessing the degree of reactivity to the agents in question and for the evaluation of the specific effects of hyposensitization. Nasal provocation tests (NPT) and bronchial provocation tests (BPT) are, however, subject to many variables (4, 22, 37). To justify their use in assessment of therapeutic efficacy, these tests have to be optimized. Even with the variables controlled as far as feasible, a challenge test is at its best only semi-quantitative. Furthermore, there exists no generally accepted standard procedure, and the reproducibility of challenge test results varies much with different procedures and population. An exact description of the method is mandatory and documentation of the range of repeatability a matter of course. Data from repeatability studies should be included in the statistical calculations. An absolute condition when using NPT and BPT for this purpose, is that the allergenic activity of the allergen preparation is the same at all stages of the challenge tests. For the comparison of tissue reactivity to different allergen types, the biologic activity of the extracts used should be known and declared in ways which make it possible to compare them (2).

Tissue reactivity to allergen challenge depends not only on the degree of immunological sensitization but also on non-specific factors (5). This is particularly important in BPT, but could also apply in NPT. The reactivity to allergen challenge should therefore preferably be correlated also to the threshold reactivity to non-specific agents such as histamine or acetylcholine. Such tests if included, should also be subject to repeatability investigations to give meaningful information.

All this makes the use of challenge tests for the assessment of efficacy a fairly demanding procedure, particularly as regards patient co-operation and motivation. Some compromise between what is optimal and what is feasible frequently becomes necessary. The degree of nasal or bronchial obstruction following quantitated challenge can be assessed by means of

objective air passage function tests (22, 37). It may be discussed which of the values in several repeated measures is the most appropriate as baseline reference. This author prefers the value which is most reproducible.

9. *Skin tests*

Skin tests may demonstrate the presence or absence of sensitization to the allergenic source represented by the extracts employed. By using great care and precision, the procedure may also show the degree of sensitization of the skin as compared with the average population allergic to the same allergenic source and as compared on each occasion. If this is the intention of the investigator, the reproducibility of the test in the hands of a particular person has to be assessed and documented. This can be done, for example, by using a histamine reference and the histamine equivalent (1 HEP) allergen concentration in octuplicate tests each time skin testing is performed with the allergen extracts during the trial period. The range of reactions obtained in the histamine reference tests should be documented in the same way as instrument calibration data are reported for technical laboratory investigation. It is not sufficient to note "prick tests were performed as described by Pepys (31)" for example. In the hands of Pepys the results with that particular method may vary less than $\pm 10\%$, but in those of the investigator in question or the nurse, possibly as much as $\pm 50\%$. Lack of reproducibility data as achieved in a particular department makes other data of skin test reactions useless.

10. *Symptom scores and drug consumption*

It is the clinical result that counts, and to the patient the results of interest are relief of symptoms and reduced need for drugs. However, these parameters may be highly subjective. They are influenced by variables the kind and number of which may be impossible for the individual patient to identify. Patients with bronchial asthma may add a component of ob-

jectivity to the symptom score by recording lung function using handy peak flow meters at given intervals, i.e. 3 or 4 times daily. For allergic rhinitis of the sneezer type of patient, an estimate of the daily number of sneezes could possibly add to the reliability of the otherwise subjective reports on symptoms and well-being (28).

If the allergenic source in question actually has been shown to be crucial to the disease, the symptoms and need for medication should correlate quite closely to the degree of exposure to the allergenic source in question. Then, information about the actual exposure experienced by the individual patient is mandatory for the evaluation of symptom scores and drug consumption. Such information may be available within certain limits for allergies to pollens and moulds. Similar data are more difficult to obtain when it concerns allergy to house dust mites and to animal danders. Exposure to other important allergens and non-allergic triggers should be kept at a minimum during the study period. If this is not possible for some of the patients, they should not be included in the trial.

Symptom score correlation to daily atmospheric pollen counts is more or less obligatory for the evaluation of efficacy of therapy for seasonal allergic rhinitis. Without information about the degree of pollen exposure, one cannot express any view as to increase or decrease of the allergic symptoms during the seasons in question. Even with atmospheric pollen counts there can be wide discrepancy in these correlations for the individual patient. The atmospheric concentration of pollen and spores at the site of the trap may differ from the concentrations found at the nose tip of the individual patient. However, with a sufficiently large number of patients within reasonable distances from the trap this problem is minimized. Correlation between symptom scores and pollen counts may be of limited value when it concerns patients with seasonal allergic asthma, unless it is shown that the asthmatic symptoms are not caused by allergy to moulds with atmospheric spore abundance more or less parallel to those of the pollen in question. Failure to present this kind of documentation reduces the readability of a number of papers reporting hyposensitization

with pollen extracts in patients with asthma occurring during the pollen season.

The "patient assessment" reported by the patient at the end of the study period is a sort of summing up of the symptom scores remembered by the patient and modulated by the attitude of the patient, for example his ability to cope with symptoms, and whether he is in sympathy with the investigator in question. This kind of yardstick for treatment efficacy is only acceptable in double-blind study designs with a large number of patients, and hardly then. This reservation also applies to the "physicians's assessment" unless an unengaged physician is responsible for this part of the study.

Drug consumption is often used as an adjunct to symptom scores.

The patients cannot be denied drugs for relief of symptoms when needed. The type of drug allowed should be clearly defined. Drugs which give quick relief from symptoms without prolonged carry-over effects to the following day(s) are to be preferred. This excludes the use of corticosteroids and makes use of antihistamines dubious.

Consumption of drugs as an effect parameter introduces variables as regards the individual attitude to taking drugs and the individual responsiveness to the given drugs.

One important question in this field is how much weight should be given to different symptom scores and drug consumption in relation to each other and in relation to other data obtained. By manipulation of the data relative to each other and with assistance from an unsuspecting statistician, the author(s) may distort the results so much that it may be concluded that almost any kind of treatment is effective – if the readers does not see through this.

11. Statistics

Although all clinical investigators ought to have some understanding of statistical principles and methods, it is recommended to seek advice and cooperation from qualified statisticians. The study design should not be finalized before both parts have reached an understanding of the problems which may be encountered. Regret-

ably, this is not always the practice. To cite Rennie (34) slightly abbreviated: "In many published papers statistical tests are improperly employed without regard to the underlying assumptions, or better tests have been ignored. Design is often defective, or the numbers tested are insufficient, or data are improperly described, selected or manipulated and the conclusions are unjustified or wrong. Schor & Karten in 1966 found such defects in the majority of 149 medical papers examined, only 28% being acceptable. Of 62 reports that appeared in the British Medical Journal during 3 months in 1976, 32 had statistical errors, and in 18 the faults were serious. Of the randomized control trial published in the New Engl. J. Med., the Lancet and other highly rated journals between 1973 and 1976, none could be rated excellent and only a third as well done".

The power of a significance test is fundamental for the evaluation of the study in question. The power is an indication of how likely it is to produce a result which is statistically significant (e.g. with $P < 0.05$ or $P < 0.01$) with a given sample size. It should be used prospectively to calculate the number of patients required for a clinical trial. Used in retrospect it may indicate how much weight could be given to the results obtained in a reported study. Power increases with sample size. Power of 80–90% is recommended. With a power of 50% or less, the study has only 50% or less chance of detecting a difference at the stated significance level.

Several authors have published graphs to assist the researcher in deciding this (7, 8–13, 17) Clark & Downie (18) among others, have published graphs for the rapid determination of the number of patients to include in each of the treatment groups in a controlled clinical trial. According to this the number of patients required to obtain a 50% chance of successful conclusion at a 5% level of significance is rather large when a new treatment is compared with control treatments previously shown to be quite effective. If the old treatment was effective in 50% and the new treatment in 75% the number required in each group is about 30. If the difference expected is from 70 to 80%, as frequently reported for treatment of allergy,

the number required is close to 200. If the better treatment improves the results by 5% from 75 to 80%, about 400 patients have to be included in the study. With 20 patients only in each group it would not be possible to detect differences in effect between old and new treatment that are less than from 15 to 40%, from 50 to 80% and from 75 to 100%!

Peto et al. have written two highly recommendable papers on statistical methods needed to design, carry out and analyse clinical trials. They too stress the need to have quite large numbers of subjects if the expected improvement is not large, even when very objective measures are available for assessment of efficacy (32, 33).

To cite Peto et al.: "Suppose there are two treatments to be compared and that the better treatment has the desired outcome ("life") in one third more than the other treatment: In other words the ratio better:worse is 2:3. The better treatment would be clearly preferable, but a trial to prove its superiority would demand an intake of hundreds of patients". And furthermore: "A large proportion (perhaps even the majority) of reports of statistically significant treatment differences in small trials are misleading, and no real differences exist. If a small trial compares a new treatment which is 20% more effective than the other treatment, then it will probably fail to reach statistical significance". A reduction of a given, defined trend (mortality) from 15 to 10% with two treatments compared, needs 1600 subjects for a power of 90% to be detected at a ($P < 0.05$) significant level.

Sample size in a controlled study depends also on the variability (precision and reproducibility) of the measurements employed as well as on the anticipated magnitude of differences. The larger the number of patients included in the trial, the better the chance to get significant results. In other words, a P -value of a small trial is less convincing than a similar P -value for a large trial, and particularly when differences between the two treatments are not expected to be extreme. The statistician should be asked which number of patients and controls should be considered the minimum under the ruling

conditions. Important information to be given for the calculation is for instance that one anticipates a placebo effect of about 30%. Furthermore, the variability (reliability, precision) of the assay methods may be a difficult point. The threshold dose for a challenge with an allergen preparation could be any dose within a 10-fold concentration of the extract simply because of the variability of the challenge method used.

If the investigation encompasses comparison of two extracts or therapeutic principles that are thought to be equally effective, the statistical calculation must be done taking into consideration that the two principles may have the same effect or that not more than a 5 to 15% difference either way can be expected. If parallels should be drawn between treatment of allergy and treatment of cancer in statistical terms, the trend "mortality" used in cancer research could be replaced by another objective trend, for instance, failure to tolerate more than a 10-fold increase in threshold dose of the allergen given in challenge tests.

With conventional preparations this failure has been about 15% in some reported studies. To show that a new preparation improves the results by 5% would then demand hundreds and hundreds of patients in both groups. In practice, many trials attempting to show small (but important) improvements like this, are deemed to be unsuccessful.

One could also present the question the other way around: With treatment A and B we anticipate a rate of success at $80 \pm 10\%$ (whereas placebo usually gives a success rate of $30 \pm 10\%$). What is the limit that can be acceptable as the range of variability (for the reaction threshold dose) in my instrumental design in order to demonstrate a significant difference ($P < 0.05$) between A and B with a power of 85% if I can include only 25 patients in each of the two groups? Distribution to patient groups and control groups should preferably be randomized. It is recommended to randomize in sub-sets of six individuals in case it should become impossible to include as many patients and controls as initially planned.

Different statistical methods may be more or less adequate for the experiments and popula-

tions in question. If one method is less favourable than another, both methods – or at least the least favourable one – should be used. If not, the referee or the reader with statistical knowledge will certainly find the less favourable way of calculation.

It is within the concept of strict scientific work not to do everything in one's power to prove that a given hypothesis is correct, but on the contrary, to try to find and evaluate arguments against the hypothesis. If the investigator himself feels that he/she is not able to provide evidence against own results, first then is there a safe enough basis for publication. In science, it is not a question of trying to market certain hypotheses or sell ideas. The reader of a scientific journal demands well-balanced, objective evidence.

All readers of scientific journals should not be considered masters of statistics, only a minority are qualified to detect fallacies in the design and statistics of a paper. Therefore, they should feel secure that this is in order when a paper has been accepted by the referees and editors of a journal. Statistical expertise should be available to the editorial staff of every high-grade medical journal (35).

12. *Immunological investigations*

It is not within the scope of this review to discuss the immunological effects of hyposensitization. The reader is referred to special reviews on this subject (3, 25). In spite of the evidence that hyposensitization exerts a beneficial clinical effect which is immunologically specific, no immunological test has yet been developed to monitor the treatment and the clinical effects of the treatment. Hopefully, competent and creative immunologists are aware of this challenge and collaborate with competent clinical investigators.

Serum and viable cells of various kinds obtained during different stages of immunotherapy in a well-conducted trial should be considered very precious gifts to dedicated researchers. Efforts should be made to have material from the patients investigated as completely as possible. The investigator should consider

his/her possibilities for studies of known and hypothetical factors regulating reagin synthesis and modulating the allergic reaction. If successful this would open up new fields of immunological knowledge. Research in the field of hyposensitization immunology up to the present time has not produced much from the clinician's point of view. New techniques and – above all – new ways of thinking are called for.

There may be many reasons for the failure to demonstrate any measureable immunological variable which can be used to monitor treatment with respect to clinical efficacy. First of all, correlation of immunological data to clinical data depends on the reliability of the latter, as discussed before. Secondly, lack of standardization and failure to include reference material makes it almost impossible to compare results obtained in different departments as presented in the published reports. The laboratory efforts in hyposensitization immunology may have been essentially fruitless because too much attention has been given to unimportant variables. Although quite a lot has been learned about immunological mechanisms and reactions following injection therapy, one gets the impression that the efforts have been wrongly directed. Authorative papers in the field may indeed hamper progress in this situation. New ideas are necessary, and may arise from information obtained through progress made in other fields of basic medicine and in allergy.

13. *Conclusion*

This reviewer has attempted to point out some of the more critical aspects of clinical trials regarding the effect of hyposensitization. Many demands must be satisfied as to the design and conductance of such a trial. Some of them are indispensable. In addition to what has been discussed are stipulations as to form of presentation including tables and graphs, balanced use of references and meticulously correct citations. Furthermore, the discussion and conclusions should give credit to objectiveness and the desire for truth rather than prestige.

Hyposensitization is the cornerstone of allergy practice (1, 30). It is a therapeutic procedure

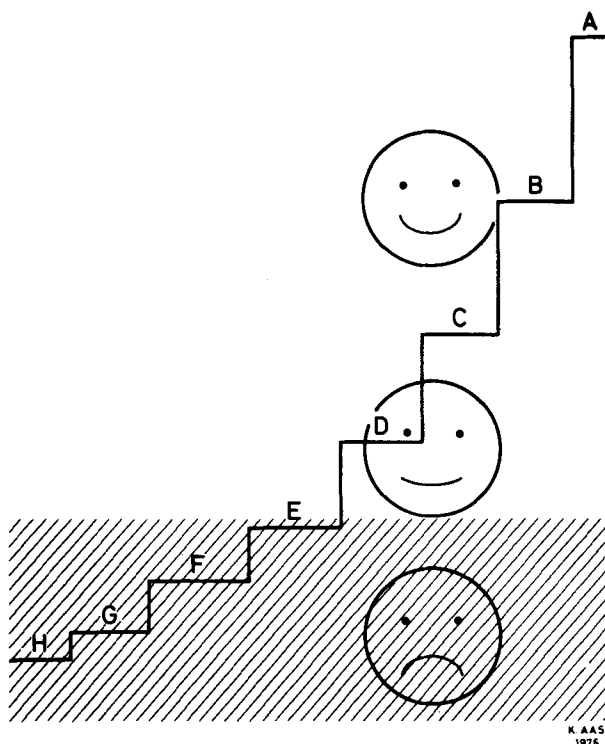


Fig. 3. The quality of papers on clinical trials may be graded following the same principle as advocated for grading of the quality of allergen extracts (1).

which is rather demanding both for the patient (and parents if the patient is a child), and for the physician. It is time-consuming and quite expensive in the long run. Nevertheless, this kind of therapy seems to be worthwhile for individuals who suffer from their allergies, provided that the therapy is optimized. Efforts should be made to elicit criteria which can identify those patients most likely to benefit from this kind of treatment.

To evaluate the effects of hyposensitization as compared with pharmacotherapy from the available literature, one has to consider the weight of evidence presented according to the quality of the investigations done, and according to the criticism exhibited in the publications in question. One should judge to what extent each study fulfils acceptable and mandatory criteria for the evaluation of causative therapy (1). The papers on this subject can be classified according to a quality scale similar to that re-

commended for classification of allergen extracts (Fig. 3 (1)).

Editors and – particularly – the reviewers have chief responsibility for filtering off all that does not meet a high scientific standard. In addition they should provide the authors with feedback impulses as and when needed. Before the paper is submitted, however, the authors should make every effort to produce a paper that can withstand the most critical scrutiny. In my judgement, what is required is the restriction of clinical trials to departments and physicians adequately trained in this discipline. This reviewer hopes that one day there will be an internationally accepted, standardized certification of clinical investigators. Then, in order to be considered for review a manuscript describing a clinical trial should be accompanied by a copy of the approved clinical investigator certificate held by at least one of the authors.

Whenever possible, the trial design should be

made in a way that facilitates comparison with other adequately performed trials. Greater use of standard procedures and well-specified material would make it possible to compile data from different centres into one meaningful pool of information (4).

The present reviewer has not identified any particular papers to illustrate the points made. It does not seem desirable to pillory particular investigators. The reader can find shortcomings by reading almost any randomly selected report on clinical trials. It may indeed be a good exercise for the reader to try to criticize such papers just as constructively as they would like reviewers to do before a paper is accepted – or not – for publication. Maybe the reader would reject the present review?

REFERENCES

- Aas, K.: Immunotherapy of bronchial asthma. In Lichtenstein, L. M. & Austen, K. F. (eds.): *Asthma: physiology, immunopharmacology and treatment*, pp. 385–400. Academic Press, New York, 1977.
- Aas, K., Backman, A., Belin, L. & Weeke, B.: Standardization of allergen extracts by appropriate methods. The combined use of skin prick testing and radioallergosorbent tests. *Allergy* 33, 130–137, 1978.
- Aas, K.: Hyposensibilisierung. Wirkung und immunologische Vorgehen. In Aas, K. & Brede, H. D. (eds.): *Allergen-Extrakte. Arbeiten aus dem Paul-Ehrlich-Institut* 73, 7–15, 1979.
- Aas, K. (ed.): Materials and methods in allergy. *Allergy* 35, 165–261, 1980.
- Aas, K.: Heterogeneity of bronchial asthma. Subpopulations or different stages of the disease? Review. *Allergy* 36, 3–14, 1981.
- Aas, K.: The optimal allergen preparation for clinical use. In Steffen, C. & Ludwig, H. (eds.): *Clinical immunology and allergology*, pp. 255–263. Elsevier/North-Holland, Amsterdam, 1981.
- Aleong, J. & Bartlett, D. E.: Improved graphs for calculating sample sizes when comparing two independent binomial distributions. *Biometrics* 35, 875–881, 1979.
- Altman, D. G.: Statistics and ethics in medical research. I. *Br. Med. J.* 281, 1182–1184, 1980.
- Altman, D. G.: Statistics and ethics in medical research. *Br. Med. J.* 281, 1267–1269, 1980.
- Altman, D. G.: Statistics and ethics in medical research. III. *Br. Med. J.* 281, 1336–1338, 1980.
- Altman, D. G.: Statistics and ethics in medical research. VI. *Br. Med. J.* 281, 1399–1401, 1980.
- Altman, D. G.: Statistics and ethics in medical research. VII. *Br. Med. J.* 281, 1612–1614, 1980.
- Altman, D. G.: Statistics and ethics in medical research. *Br. Med. J.* 282, 44–47, 1981.
- Bluestein, J.: On children and proxy consent. *T. Med. Ethics* 4, 138–140, 1978.
- Baer, H., Godfrey, H., Maloney, C. J., Norman, P. S. & Lichtenstein, L. M.: The potency and antigen E content of commercially prepared ragweed extracts. *J. Allergy* 45, 347–354, 1970.
- Bernstein, J. L., Perera, M., Gallagher, J., Michall, J. G. & Johansson, S. G. O.: In vitro cross-allergenicity of major aeroallergenic pollens by the radioallergosorbent technique. *J. Allergy Clin. Immunol.* 57, 141–152, 1976.
- Boag, J. W., Haybittle, J. L., Fowler, J. F. & Emery, E. W.: The number of patients required in a clinical trial. *Br. J. Radiol.* 30, 371–381, 1979.
- Clark, C. J. & Downie, C. C.: A method for the rapid determination of the number of patients to include in a controlled trial. *Lancet* ii, 1357–1358, 1966.
- Cockburn, F., Dudgeon, J. A., Gairdner, D. M. T. & Jackson, A. D. M.: Guidelines to aid ethical committees considering research involving children. *Arch. Dis. Child.* 55, 75–77, 1980.
- Declaration of Helsinki, *WHO Chronicle* 30, 360–362, 1976.
- Editorial: Controlled trials: Planned deception? *Lancet* i, 534–535, 1979.
- Engström, I.: The bronchial provocation test and criteria for positive reaction. *Allergy* 35, 256–258, 1980.
- Frostad, A. B., Grimmer, Ø., Sandvik, L. & Aas, K.: Hyposensitization. Comparing a purified (refined) allergen preparation and a crude aqueous extract from timothy pollen. *Allergy* 35, 81–95, 1980.
- Hunt, K. J., Sobotka, A. K., Valentine, M. D., Yunginger, J. W. & Lichtenstein, L. M.: Sensitization following Hymenoptera whole body extract therapy. *J. Allergy Clin. Immunol.* 61, 48–53, 1978.
- Ishizaka, K. & Ishizaka, T.: Mechanisms of reaginic hypersensitivity and immunotherapy. *Lung* 155, 3–22, 1978.
- Levine, K. J. & Lebacqz, K.: Ethical considerations in clinical trials. *Clin. Pharmacol. Ther.* 25, 728–741, 1979.
- Lichtenstein, L. M.: An evaluation of the role of immunotherapy in asthma. *Am. Rev. Respir. Dis.* 117, 191–197, 1978.
- Mygind, N.: Clinical investigations of allergic rhinitis and allied conditions. *Allergy* 34, 195–208, 1979.
- Norman, P. S.: A review of immunotherapy. *Allergy* 33, 62–70, 1978.
- Patterson, R.: Clinical efficacy of allergen immunotherapy. *J. Allergy Clin. Immunol.* 64, 155–158, 1979.
- Pepys, J.: Skin tests in diagnosis. In Gell, P. G. H., Coombs, R. R. A. & Lachman, P. J. (eds.): *Clinical aspects of immunology*, pp. 55. Blackwell, Oxford, 1975.
- Peto, R., Pike, M. C., Armitage, P. et al.: Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design. *Br. J. Cancer* 34, 585–612, 1976.
- Peto, R., Pike, M. C., Armitage, P. et al.: Design and analysis of randomized clinical trials requiring prolonged observation of each patient. II. Analysis and examples. *Br. J. Cancer* 35, 1–39, 1977.
- Rennie, D.: Vive la différence ($p < 0.05$). *New Engl. J.*

- Med. 299, 828-829, 1978.
35. Shuster, J. J., Binion, J., Moxley, J. et al.: Statistical review process. Recommended procedures for biomedical research articles. *J. Am. Med. Assoc.* 235, 534-535, 1976.
 36. Simmons, B.: Problems in deceptive medical procedures: An ethical and legal analysis of the administration of placebos. *J. Med. Ethics.* 4, 172-181, 1978.
 37. Taylor, G., MacNeil, A. R. & Freed, D. L. J.: Assessing degree of nasal potency by measuring peak expiratory flow rate through the nose. *J. Allergy Clin. Immunol.* 52, 193-198, 1977.
 38. Taylor, W. T., Ohman, K. L. & Lowell, F. C.: Immunotherapy in cat-induced asthma. Double-blind trial with evaluation of bronchial responses to cat allergen and histamine. *J. Allergy Clin. Immunol.* 61, 283-287, 1978.
 39. Turkeltaub, P. C., Marsch, D. G., Lichtenstein, L. M. & Norman, P. S.: Development of long-lasting immediate hypersensitivity in non-atopic volunteers parenterally immunized with a purified grass pollen extract. *J. Allergy Clin. Immunol. (Abstracts)*, 61, 171, 1978.

Address:

K. Aas
The Allergy Unit
Pediatric Research Institute
Rikshospital
Oslo 1
Norway