

Basic principles in the planning of clinical trials in surgical oncology

W. Draxler¹ and M. Mittlböck²

¹Department of Surgery, Medical University of Vienna, and ²Core Unit for Medical Statistics and Informatics, Medical University of Vienna, Vienna, Austria

Received December 2, 2005; accepted after revision December 20, 2005
© Springer-Verlag 2006

Grundlagen von Studienplanung in der chirurgischen Onkologie

Zusammenfassung. *Grundlagen:* ICH (Internationale Konferenz zur Harmonisierung der technischen Anforderungen für die Registrierung von Pharmazeutika) stellt Richtlinien bezüglich der Implementierung von klinischen Studien zur Verfügung. Es ist für alle Beteiligten an einer klinischen Studie verpflichtend, diese Richtlinien im Rahmen der „Good Clinical Practice“ einzuhalten.

Methodik: Die Hauptmerkmale einer klinischen Studie beinhalten die im Folgenden beschriebenen Punkte: Hintergrund und allgemeine Ziele, spezielle Ziele, Patientenauswahlkriterien, Medikationsverabreichung, Evaluierungsmethoden, Studiendesign, Registrierung und Randomisierung, Patienteneinverständnis, benötigter Stichprobenumfang, Monitoring des Studienfortschritts, Dateneingabebögen und Datenbearbeitung, Protokollabweichungen, Planung statistischer Analysen und administrative Verantwortlichkeiten.

Ergebnisse: Alle oben angeführten Aspekte sollen schon in der Planungsphase einer klinischen Studie diskutiert und im Studienprotokoll festgehalten werden. Das Studienprotokoll stellt eine Leitlinie für alle an der Studie beteiligten Personen dar.

Schlussfolgerungen: Besonders wichtig für den Erfolg einer klinischen Studie ist eine möglichst klare und exakte Formulierung der Forschungshypothesen und die sorgfältige Auswahl von primären und sekundären Zielgrößen.

Schlüsselwörter: Leitlinien, Studiendesign, Randomisierung, Statistische Analyse, Stichprobenplanung.

Summary. *Background:* ICH (International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use) provides guidelines on the implementation of clinical trials. All study

participants are obliged to follow these guidelines in line with “Good Clinical Practice”.

Methods: The main features of a clinical study include the following items: Background and general aims, specific objectives, patient selection criteria, treatment schedules, methods of patient evaluation, trial design, registration and randomization of patients, patient consent, required size of study, monitoring of trial progress, forms and data handling, protocol deviations, plans for statistical analysis and administrative responsibilities.

Results: All items mentioned above should already be discussed in the planning stage of a clinical trial and addressed in the study protocol. The study protocol provides a guideline for any person involved in the trial.

Conclusions: For the success of a clinical trial, it is especially important to have a clear and exact definition of the study hypotheses and to choose primary and secondary endpoints very carefully.

Key words: Guidelines, trial design, randomization, statistical analysis, sample size determination.

Introduction

The design of a clinical trial, from initial rather vague ideas about treatment innovation to a detailed plan of action, is mostly a complicated process.

We distinguish between different types of trials. *Phase III*-trials undertake a full-scale evaluation of investigated treatments. After a drug has proved to be effective (*Phase II*), the next step is its comparison with the current standard treatment(s) for the same condition in a large trial involving a substantial number of patients. *Phase II*-Trials deal with initial clinical investigation for treatment effect, while *Phase I*-Trials investigate clinical pharmacology and toxicity.

There are three fundamentals of trial design which must be defined accurately already at an early stage:

- which patients are eligible
- which treatments are to be evaluated and
- how each patient’s response is to be assessed.

Correspondence: Wolfgang Draxler, MSc, Department of Surgery, Medical University of Vienna, Spitalgasse 23, 1090 Vienna, Austria.

Fax: ++43-1-404006918

E-mail: wolfgang.draxler@meduniwien.ac.at

These aspects are part of the study protocol which contains a specification how the trial is to be conducted. The protocol provides detailed information on the trial procedure relating to each individual patient. Hence, the trial requirements for patient entry, treatment and evaluation plus data collection procedures need to be exactly stipulated. Besides, the trial's motivational background, specific aims and the rationale behind the chosen study design are stated in the protocol. It is essential to formulate such a document so that everyone involved in the proposed trial is fully informed.

According to Pocock [1] the *main features of a study* can be summarized in a list of 14 items, which indeed should be addressed in the study protocol:

1. Background and general aims
2. Specific objectives
3. Patient selection criteria
4. Treatment schedules
5. Methods of patient evaluation
6. Trial design
7. Registration and randomization of patients
8. Patient consent
9. Required size of study
10. Monitoring of trial progress
11. Forms and data handling
12. Protocol deviations
13. Plans for statistical analysis
14. Administrative responsibilities

In the following these fourteen points are described shortly. Further details about planning and conducting clinical studies can be found in Pocock [1], Hulley et al. [2], Schumacher and Schulgen [3] and Shapiro and Louis [4].

Principles

Background and general aims

A description of the background and general aims of the trial is a useful preliminary, which helps to explain why the trial is considered worthwhile and how it builds on experience gained from previous research.

Specific objectives

Hypotheses regarding treatment efficacy and safety which are examined by the trial have to be precisely defined by specific objectives of the trial. They are a more concise worded definition of the investigated hypotheses.

Patient selection criteria

Any clinical trial requires a precise definition of the patients' eligibility for inclusion. One essential aim is to ensure that the selected patients included in the trial are representative for the future group of patients to whom the trial findings should possibly be applied to. One must not be too restrictive about patient entry as otherwise the trial could remain small and the results lack generality.

Treatment schedules

Drug regimes must be defined in the study protocol. The following features should be addressed: Drug formulation, route of administration (e.g. oral, intravenous,

intramuscular), amount and frequency of each dose (determined from experience in Phase I/II trials), duration of therapy (fixed or dependent on each patient's progress), side-effects, dose modification and withdrawal, patient compliance with therapy, ancillary treatment and patient care, packaging and distribution of drugs (particularly important in the case of multi-centre or double-blind studies) and comparison of treatment policies.

Methods of patient evaluation

In a trial therapy evaluation of the patient's progress has to be evaluated in an objective, accurate and consistent manner. Methods for assessing outcome need to be precisely defined in the study protocol (appropriate time interval between evaluations) and it is clearly not sufficient to control patient's progress as normally in general clinical practice. Routine case notes are usually far too vague, inconsistent and subjective.

In surgical oncology typical trials involve the following criteria of response:

1. survival time (time from randomization until death)
2. recurrence free survival time (time from randomization until first occurrence of local recurrence, distant metastases or contralateral cancer)
3. achievement and duration of tumor response (partial or complete reduction in tumor size)
4. change in performance status
5. occurrence of toxicity
6. occurrence of side-effects

These criteria are often divided into primary and secondary target variables also called endpoints. The primary endpoint describes the variable of highest interest and is used to answer the main question of the trial. Secondary target variables include all other criteria of interest investigated in the study and give additional information. A typical example in surgical oncology would be: Primary endpoint is recurrence-free survival secondary endpoints are overall Survival (death of any cause) and occurrence of predefined toxicities or side-effects.

Trial Design

In most cases trial designs in oncology are conducted to compare treatments between parallel groups. It is the rare exception that one group consists of untreated controls as this can only be ethical if no effective therapy exists. Trials including a placebo group are just as seldom. The prime reason for introducing placebo controls is often to make patient attitudes to the trial as similar as possible in treated and untreated groups. Another type of study design is a crossover trial. Such trials can only be performed for chronic and non-curable diseases. In their simplest form each patient receives two treatments one after the other, where the order of administered treatments being randomized.

The experimental units in clinical trials are patients. The idea of a random assignment to a particular treatment is not intuitively appealing. In a certain sense a randomized comparison appears contrary to the efforts of the clinician to give every patient the best possible care and it implies a loss of freedom for both patient and clinician. So, why should randomization be considered such a key

issue in the conduct of clinical trials? The reason for this is straightforward. All alternative approaches are indicated by some deficiencies. In a non-randomized trial the clinician would tend to give the new treatment a reasonable chance of success by selecting less seriously ill patients. Consequently, such a selected experimental group of patients will appear to have surprisingly good outcome in comparison to the general routine.

To sum up, one can say that randomized controlled trials are an essential tool for testing the efficacy of therapeutic innovations. A well performed randomization process guarantees that there is no bias in the selection of patients for the different treatments. But randomization alone cannot rule out that the comparison of treatments may be distorted if the patient himself and those responsible for treatment and evaluation know which treatment is being used. Replying to this the term *Blinding* becomes meaningful. In a double-blind trial neither the patient, nor physician nor an evaluator is aware of the actual treatment of patients. The patient blinding avoids any “treatment effect” which is psychologically justified. If the patient knows he gets the new (“better”) treatment then he may particularly benefit by positive thinking. Blinded physicians will not tend to observe patients on the new treatment more closely than the progress of others on standard therapy. Furthermore, decisions on dose modification, intensity of patient examination cannot be influenced by a blinded physician. A blinded evaluator assessing patients outcome should be able to act as objective as possible.

The decision for a blinded trial depends on certain circumstances. A careful consideration of the following points should help to decide in each case:

1. Ethics: A double-blinded procedure should not result in any harm to a patient (e.g. quick unblinding in case of severe side-effects).
2. Practicality: For some treatments it would be completely impossible to arrange a double-blind trial. (e.g. comparison of surgery versus conventional therapy)
3. Avoidance of bias: One needs to assess how serious the bias might be without blinding.
4. Compromise: Sometimes even partial blinding can be sufficient to reduce bias in treatment comparison or only blinded evaluations are sufficient

Registration and randomization of patients

For any patient who might be considered suitable for a trial a formal sequence of events should take place: (1) patient requires treatment, (2) checking eligibility, (3) clinician willing to accept randomization, (4) patient consent, (5) formal entry on trial, (6) treatment assignment with randomization procedure, (7) on-study forms completed, (8) treatment commences.

As mentioned in the preceding item randomization is an essential necessity for a fair and unbiased comparison between treatments. Frequent randomization methods are:

1. Simple randomization: Principle of tossing a coin. This method is very simple and completely unpredictable but doesn't allow for stratification (balancing of other important prognostic factors).

2. Random permuted blocks within strata: Ensures exactly equal treatment numbers at certain points (depending on block size) in the sequence of patient assignments within strata. However there is a risk of predictability and unbalances may arise in case of many strata. Thus it is only feasible for a relatively small number of strata.
3. Biased coin: The probability for each treatment depends on all preceding randomization results in such a way that the treatment with currently fewer observations has a higher probability of being selected for the next randomization. The procedure rather prevents unbalanced treatment groups, but can only be performed by computers.
4. Minimization: This method is similar to random permuted blocks but it is also practicable for a higher number of strata. The purpose of this method is to balance the marginal treatment totals for each level of the considered prognostic factors. This method is widely used and computer programs are available, e.g. internet randomization. For more details I refer to the references at the end of this article.

Patient consent

The declaration of Helsinki and ICH-Guideline E6 [6] states that each potential subject must be adequately informed of the aims, methods, anticipated benefits and potential hazards of the study and the discomfort it may entail. The patient should be informed that he is at liberty to abstain from participation in the study and that he is free to withdraw his consent to participation at any time. Prior to a patient's participation in a trial, the written informed consent form should be signed and personally dated by the patient.

Required size of study: Sample size calculation

Sample size calculation is necessary to ensure that a truly existing difference for the primary endpoint between treatment arms will be detected with high probability (also called power, typically 80 to 90 %) and to keep the false-positive error (Type I or α -error) small, usually at 5 %. As the term “truly existing difference” is very unspecific, the sample size calculation is based on the minimum clinically relevant difference between groups which has to be defined in the context of the underlying medical situation and medical consequences. E.g. what improvement has to be achieved to change to the new treatment in routine clinical situations?

Usually two hypotheses are set up: the null hypothesis (H_0) assumes no difference between arms whereas the alternative hypothesis (H_1) states that there is a difference, this is a typical situation for two-sided tests. In rare cases it may make sense that only deviations in one direction are of interest, then e.g. H_1 states that the new treatment is superior to standard treatment. In such one-sided situations it would have the same consequences if either the therapy groups are rather comparable or if the new therapy turns out to be worse than standard therapy. However such one-sided tests have to be unambiguously justified in the study protocol.

8.1.16 Anzahl der Gruppen:	2	
8.1.17 Stratifizierung:	<input type="checkbox"/> nein <input checked="" type="checkbox"/> ja: Kriterien:	tumor stage, lymph node stage, grading
8.1.18 Messwiederholungen:	<input checked="" type="checkbox"/> nein <input type="checkbox"/> ja: Zeitpunkte:	
8.1.19 Hauptzielgröße:	overall survival	
8.1.20 Nullhypothese(n):	no difference in overall survival between both treatment groups	
8.1.21 Alternativhypothese(n):	improvement of survival proportion at 5 years from 60% under standard therapy to at least 70% under new adjuvant therapy	
8.1.22 Nebenzielgrößen:	recurrence-free survival, adverse events	
<hr/>		
8.2 Studienplanung		
Die Fallzahlberechnung basiert auf (Alpha = Fehler 1. Art, Power = 1 – Beta = 1 – Fehler 2. Art):		
8.2.1 Alpha:	0,05	8.2.2 Power: 0,85 8.2.3 Stat.Verfahren: test based on exponential survival
<hr/>		
8.2.4 Multiples Testen:	<input checked="" type="checkbox"/> nein <input type="checkbox"/> ja: Korrekturverfahren:	
8.2.5 Erwartete Anzahl von Studienabbruchern (Drop-out-Quote):	0,05 expon. drop-out	
<hr/>		
8.3 Geplante statistische Analyse		
Population:	<input checked="" type="checkbox"/> 8.3.1 Intention-to-treat	<input type="checkbox"/> 8.3.2 Per Protocol
8.3.3 Zwischenauswertung:	<input checked="" type="checkbox"/> nein <input type="checkbox"/> ja: Abbruchkriterien:	
8.3.4 Geplante statistische Verfahren:	Kaplan-Meier graphs, log-rank test, proportional hazards regression model (Cox)	

Fig. 1. Selected sections from the submission form to ethics committees in Austria

Based on the result of the so-called test-statistic (determined from the study data) either H_0 is rejected and H_1 accepted and a statistically significant difference is assumed or H_0 can not be rejected. Only with proper sample size and consequently high power it can be stated that H_0 is accepted if the test is not significant.

In the following, three typical sample size calculation scenarios in oncology are described, differing by their underlying response type, where only situations with two treatment arms and equal sample sizes are discussed here. For all sample size calculations a one- or (preferably) two-sided test procedure, an α -error (usually 5%) and the power of the study (usually 80 to 90%) has to be chosen. All other parameters depend on the response type of the primary endpoint:

1. *Survival studies (time-to-event outcome, often recurrence-free or overall survival)*: Assuming constant hazard rates of an event over time, the survival proportion at a specified time-point under standard therapy should be known (from other studies or from the own expert knowledge). Then the minimum clinically relevant improvement of the survival proportion under the experimental therapy has to be defined. This is equivalent to define the difference between groups as a hazard-ratio. From this information the number of events needed for the analysis of the study can be calculated. The total number of patients needed to observe this number of events depends on the accrual rate (number of patients to be recruited over a specified time period) and the length of the accrual-period (duration of recruitment) and the follow-up-period (period from end of recruitment to analysis). E.g. for a fixed accrual rate a longer follow-up period decreases the total number of patients needed (the accrual period can be shortened) but the total length of the study would increase. If drop-outs are likely, the expected drop-out rate should be considered.

Example: If the efficacy of a new adjuvant therapy in colorectal cancer should be assessed, it seems to be

plausible that 400 patients fulfilling all inclusion and exclusion criteria can be recruited to the study per year. Furthermore a drop-out rate of 0.05 is assumed. If the primary end point is overall survival and the survival proportion at 5 years is 60 % under standard treatment and the new adjuvant therapy should improve the overall survival at least to 70 %, then the following result can be calculated using sample size software: 278 events have to be observed for the final analysis to detect this 5-year difference in survival, equal to a hazard ratio of 1.432, with a two-sided test, a significance level of 5 % and a power of 85 %. This can be achieved e.g. by recruiting 600 patients per group within three years and a further follow-up period of 2 years (total study duration of 5 years) or by recruiting 500 patients per group within 2.5 years and a further follow-up of 3.2 years (total study duration of 5.7 years). The corresponding sections (8.1.16-8.3.4) from the submission form to ethic committees in Austria are shown in Figure 1.

2. *“proportion of success” (binary outcome)*: The success rate under standard treatment should be known and the minimum clinically relevant change in the success rate under experimental therapy should be specified.

Example: If an increase from 3 cycles of neoadjuvant therapy to 6 cycles in breast cancer patients improves the rate of complete pathological remissions from 7 % to at least 18 %, then this difference can be detected by a Chi-square test with a power of 85 % and a significance level of 5 % (two-sided) if 161 patients are included in each group. Assuming a drop out rate of 5%, at least 170 patients per group have to be enrolled in the study.

3. *normally distributed continuous outcome*: The smallest difference of the two means, which is clinically relevant, and the common standard deviation has to be specified.

Example: Bone mineral density of breast cancer patients decreases to a mean value of 0.85 g/cm² after three years of hormone therapy. An additional bisphosphonate therapy should increase the mean bone mineral density after 3 years to at least 0.9 g/cm². It is known that the standard deviation for bone mineral density measurements of patients with hormone therapy alone is 0.1 and it is assumed that the standard deviation in the hormone+bisphosphonate therapy group is similar. To detect a difference in means of 0.05 g/cm² by a two sided t-test with a power of 85 % and a significance level of 5 %, 73 patients per group are needed. Assuming a drop out rate of 5%, at least 77 patients per group have to be enrolled in the study.

In the case of other scenarios such as comparisons between more than two groups, unequal sample sizes between groups, equivalence studies, cross-over designs or outcomes of special structure different sample size calculation methods (not discussed here) have to be applied.

All sample size calculations here have been carried out using nQuery Advisor 6.0 [5].

Monitoring of trial progress

In most clinical trials patients' responses to treatment are observed sequentially. A patient is expected to have his "status" assessed at regular intervals as predefined in the study protocol. Why is this monitoring over time essential? One reason is to check protocol compliance. If early results indicate difficulties regarding compliance alterations to the protocol would be inevitable. Continuous monitoring of trial progress also allows reporting of side-effects, particularly severe toxic reactions to a new therapy and thus permits instantaneous reactions, e.g. dose modification. A permanent data monitoring and recordation process also allows to keep track of any inconsistencies in the data and their elimination, as checks can be carried out during progression of the study.

Forms and data handling

It has to be defined in the study protocol which data have to be collected for each patient. On the basis of this definition case record forms (CRF) need to be designed. In this context it is important that forms are clearly arranged and self-explanatory and data are collected only if truly of interest. Forms have to be designed and distributed before the trial starts so that there is no delay in data recording.

All data have to be collected, checked and organized in an efficient, adequate and feasible way. These activities can be summarized by the term "Data management" which becomes more and more important as quality demands on clinical studies have substantially increased over the last decades.

Protocol deviations

It is beyond doubt that the main target is to avoid any protocol deviations as they may bias the therapeutic comparison. Unfortunately, even though one tries there may be some deviations anyway. Priority should be given to the reduction of the number of ineligible patients. A clear definition of patient eligibility should be included in the

protocol and each investigator should be encouraged to run through this check-list carefully every time he is about to enter a patient. Further sources of protocol deviations may arise after a patient has already entered a trial. This could be non-compliance, incomplete evaluation or withdrawal. All these aspects have to be considered when statistical analyses are performed. It is of immense importance to evaluate the influence of protocol deviations with respect to the results of analyses and to check if study results are robust against them.

Plans for statistical analysis

There are basically three types of response data in clinical trials:

1. Qualitative response: each patient is classified into one of several response categories according to predefined evaluation criteria. In the simplest form, one can have two possible outcomes, success or failure.
2. Quantitative response: If a quantitative measure of response exists it is usually best to use its actual numerical value in the analysis and not to categorize it; this allows a more precise evaluation with respect to the size of differences in outcome.
3. Time-to-event: This type of response is most common in clinical oncology. In this case the main evaluation of therapy is carried out in terms of the time to some major event.

Another quite important term in this context is "prognostic factors". These may be personal characteristics (e.g. age, sex) and other on-study baseline data if they are observed before the treatment starts. The main objective is to determine such factors which may have a significant influence on patient's response. Multiple regression models can assess them together and may detect interactions.

All statistical analyses should be carried out for at least two different underlying patient populations. The Intention-to-treat analysis is usually based on all randomized patients independent if all inclusion criteria have been fulfilled or if the treatment has been applied as defined in the study protocol. The Per-protocol patient population only includes patients that fulfil the protocol in a predefined extend. A Per-protocol analysis is based on this dataset. It is quite obvious that results of these two types of analyses have to be compared carefully and emerging deviations have to be examined as systematic dropouts may have an impact on the results.

Generally, all data should firstly be described descriptively if applicable. Dependent on the type of response there are various possibilities of analysing the data, reaching from simple two-group comparisons with a t-test over ANOVA (Analysis of Variance), corresponding non-parametric methods, linear regression, Kaplan-Meier log-rank-test to univariate and multiple logistic or Cox proportional hazards regressions models. Concerning the above it is demanded to stipulate all statistical analyses already before the trial starts. These aspects have to be addressed in the study protocol at least, but if large-scale explanations and definitions would be necessary then a separate Statistical Analysis Plan suits this purpose best.

The decisions about the choice of response variables and analyses strategies have to be taken very carefully as this directly influences the potential of a successful trial.

Interim Analyses

If interim analyses are intended, type and number have to be defined in the study protocol. In this regard it is essential to deal with the problem of multiple testing. Interim analyses have to be performed at a lower significance level to ensure that the overall, α -error is maintained.

Interim analyses may possibly allow for earlier study termination in case of strong differences between treatment groups that are detected by such early-stage analyses.

Administrative responsibilities

Naturally, any trial benefits from clearly defined leadership, usually by an experienced principal clinical investigator. Any trial requires a coordinating centre to handle all administrative matters and a responsible person with statistical experience, who should be involved from the beginning of the study planning phase to the final analysis of the data. In general, the success of a trial relies heavily on each individual participant being fully informed and able to carry out his responsibility.

ICH/GCP

ICH ("International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use") [6] is a unique project that brings together the regulatory authorities of Europe, Japan and the United States and experts from the pharmaceutical industry in the three regions to discuss scientific and technical aspects of product registration. This conference has agreed on guidelines on four topics regarding quality, safety, efficacy and multidisciplinary aspects. Most important from a clinical oncology point of view are the subcategories "Good Clinical Practice" (E6) and "Clinical Trials" (E7–E11) that are part of the efficacy topic.

All basic principles described above are pointed out in detail in a corresponding ICH-guideline.

Current performance of clinical studies is affected by these ICH-Guidelines. It is desired that clinical trials have to be carried out in consideration of these guidelines and anyone being involved is desired to act upon ICH. By reference to ICH E6 the guidelines of Good Clinical Practice are also legally fixed in the Austrian Law of Drugs (AMG) as of May 1st, 2004.

Recommendations: Take home messages

1. Take statistical issues serious before starting a trial because statistical analysis of results, no matter how cleverly done, can never rescue a poorly designed study.
2. Choose your primary and secondary endpoints carefully, define your study hypotheses as exactly as possible and be realistic with respect to realizable recruiting rates
3. Allocate your patients to treatment arms by proper randomization to avoid bias
4. Define and check inclusion and exclusion criteria carefully and don't enter a patient into a trial on suspicion
5. Besides the p-value it is also fundamental to describe the extent of the difference between groups (effect size) and to check if it is really clinically relevant.

References

1. Pocock SJ (1984) *Clinical Trials – A Practical Approach*. John Wiley & Sons, Chichester
2. Hulley SB, Cummings SR, Browner WS, Grady D, Hearst N, Newman TB (2001) *Designing Clinical Research*. Second Edition. Lippincott Williams & Wilkins, Philadelphia
3. Schumacher M, Schulgen G (2002) *Methodik klinischer Studien – Methodische Grundlagen der Planung, Durchführung und Auswertung*. Springer, Berlin
4. Shapiro SH, Louis TA (1983) *Clinical Trials – Issues and Approaches*. Marcel Dekker, Inc. New York
5. nQuery Advisor® 6.0, Copyright © 1995-2005, Janet D. Elashoff
6. Official web site of ICH, <http://www.ich.org>, April 2005