Sonderdruck aus

Advances in Clinical Pharmacology Vol. 13

Problems of Clinical Pharmacology in Therapeutic Research: Phase I

Proceedings of the 14th International Symposium on Clinical Pharmacology, Tokyo, July 1976

Volume Editors

H. P. Kuemmerle/T. K. Shibuya/E. Kimura

Printed in Germany.

Urban & Schwarzenberg · München-Wien-Baltimore 1977

Methodological Considerations for Phase I Studies: System Analysis, Experimental Design, Statistical Evaluation and Data Management

Karl K. Überla
Professor of Medical Information Processing
Statistics & Biomathematics
University of Munich
Marchioninistr. 15, D-8000 Munich 70
Germany (GFR)

Phase I studies, conducted in a formal way, are not yet old. The present state of affairs evolved partly by systematic considerations, partly by chance and success. The legal requirements in the different parts of the world are different and have had their impact on the present way of conducting clinical trials with a new drug for the first time in man. The result of the various influences—scientific, clinical industrial, legal, historical, chance—is a compromise, which could most likely be improved.

So I would like to take a fresh look at phase I studies from the point of view of general methodology. The general system approach has to my knowledge not yet been applied to this area. Experimental design, statistical evaluation, and data management by computers are methodological parts of the general system approach as applied to phase I studies.

I shall omit the legal regulations, which are different in different countries. I shall also omit the preclinical and technical requirements and I shall not go too much into details, since I think the simple and general considerations are more important. The details of clinical trials are well known, so it is not necessary to mention them. I shall describe the major parts of phase I studies in a general and formal way, show some bottlenecks and problems suggested by this description, and make comments and proposals for some problems. I hope to show some different points of view concerning the problems of phase I studies, although their solution is beyond the scope of this paper.

1. DESCRIPTION OF PHASE I SYSTEMS

Phase I trials can be described as the planned and organized application of a new substance or drug to man, for the first time and for the first few

human cases. What are the goals to be achieved by such trials? The description and specification of the goals is the first step of our analysis.

1.1 System of goals.

There is no single and simple goal for a phase I system. There are at least five different groups of persons or points of view with different goals:

The *consumers* want:

—to get effective or more effective drugs
—to get such drugs for the most frequent and most important diseases and for as many other illnesses as possible
—to get such drugs as quickly as possible (for current patients and for future generations)
—to avoid undesirable side effects.

Important points for consideration in respect to these goals are the frequency of the diseases, the prognosis of the diseases, and the results of the present treatments.

The *subjects* have a different set of goals. They want:

—to get some benefit from the trial (better treatment, money, other advantages)
—not to be impaired by the trial (discomfort and pain during trial, possible drug effects, total time spent)
—not to lose basic rights (informed consent).

Some of the goals of *pharmaceutical companies* are:

—to prove tolerance and effectiveness
—to fulfill the legal and scientific requirements for phase II
—not to spend too much money.

The *investigator* (physician, clinical pharmacologist) has as his goals:

—to find a new and effective drug, which helps better than other drugs for a specified illness
—to be successful as a scientist and to gain a reputation within the scientific community
—to earn some money for doing the job.

Finally the goals from a *scientific* point of view are:

—to prove tolerance
—to show effectiveness (pharmacodynamic effects)
—to establish dose range

—to monitor any toxic or side effects
—to study pharmacokinetic effects
—to study biovailability.

The goals of the different participants in phase I trials are not the same; they are even conflicting in some respects. The goal of the consumer, to get effective new drugs as quickly as possible, might be in conflict with the goal of the subject not to be impaired or not to lose basic rights. The goal of the company not to spend too much money might be in conflict with the goal of the investigator to be successful as a scientist, and so on. There is a rather complex set of goals; the single goals should be achieved at least partly or in various combinations during phase I.

The specific goals for every trial must be set clearly. Part of the difficulties and the frustration of phase I studies originates from the fact that the goals are conflicting, not sufficiently specified, and the relative weights are not set beforehand.

Now let us look at the individual trial as a system in itself.

1.2 The individual trial as a system

Basically, the investigator examines the subject, takes measurements of various kinds, applies physical tests, and gets back some information. This dynamic relationship is established before the application of a new drug. After the application of the new drug the difference or the change in symptoms and signs or the change in measurements is the desired information for the investigator. In case this information shows some dangerous or toxic effects, the investigator stops the drug application and applies some therapy.

The relation between subject and investigator is a rather complex one. The set of measurements and obtainable information have to be specified, including the exact time course. The drug must be applied in the proper dose and at the proper time and the therapy for possible toxic effects has to be available. The investigator has also to convince the subject that the danger is reasonable and that the benefit of the trial will be greater than the possible hazards.

An individual phase I trial is usually not confined to one patient. Several patients—one after the other or in groups—are taken through the trial.

There are the following *basic steps in conducting* a phase I trial:

1. The *goals* of the trial must be defined. The questions here are:
   a) What is the mix of the goals in this special trial?
   b) Are the goals sufficiently specified?
2. The *trial* must be *defined* in all details. The question here is: What is the optimal design for the defined goals? Often the goals are changed during the process of detailed design and definition of the trial.

starting phase II are set broadly, one will soon begin phase II and if they are set narrowly the opposite will be the case. Similar statements are true for the conditions under which the next trial is started.

The evaluation of accumulated evidence after each individual trial is a very decisive step in this process. The conditions for the following decisions might be changed at this point, the goals for the next individual trial might be set and the performance of the trial might be measured in some way.

So there is a general flow chart for phase I trials. Phase I is a dynamic system of consecutive individual trials. During the course of phase I the goals and the stop- and go-conditions are adapted and changed.

One of the major problems with such a system is the consecutive mixing of goals for the individual trials. Usually one starts with tolerance and tries to prove the pharmacodynamic effects. During later stages by increasing doses the dose range is established, and pharmacokinetic effects (absorption, distribution, metabolism, excretion) and bioavailability (half-life time and elimination rate) are of interest.

Generally one could give all possible weights to the goals and define the individual trials according to their goals. It would, for instance, be possible to start from the very beginning with the study of pharmacokinetics and/or bioavailability. So far only Ia and Ib trials are differentiated but the definition is not quite consistent in the literature. The question remains, what is the optimal mix of goals for each individual trial, which can be answered only after evaluating the specific evidence. To some extent in most trials every goal is investigated or information is gathered for its later investigation.

The definition of types of phase I trials can be further developed according to the spectrum of the goals.

We now have a formal description of the phase I system and its parts and can start with the consideration of some problems and bottlenecks inherent in this system and we can try to develop some proposals.

2. SOME PROBLEMS, BOTTLENECKS, AND PROPOSALS

2.1 Conflicting goals/medical balance sheet

The priority for the various goals can be set in a different way depending largely on the culture and the feeling of the living population at the time. In our time it seems to be appropriate to give the goals of the patients and the goals of the subjects the highest priority.

This would mean weighing the goals and the interests of the patients versus the goals and the interests of the subjects in some way. This is of

course a highly subjective area and measurement will not be easy. However one could reach a rather general conclusion without establishing all the details.

Imagine we could establish a *medical balance sheet* for phase I trials. The risk for the subjects could be measured in the number of life-months, which are possibly impaired or even not lived, in case the subjects die. The benefit for the possibly treated patients could be measured in the same unit, namely number of life-months, which are lived in a better way or which are even gained, in case treated patients live longer. So the unit of measurement would not be dollars, but the individual life-month gained or lost, impaired or improved. These units of measurement could be summed up for all subjects and all treated patients.

Since there is always a fixed number of subjects in phase I, the possible risk will always be a constant. These parallel lines might be higher or lower depending on the weight one gives to the life-months of subjects and on the estimated probability that there will be an impairment by the trial. But the summed risk will always be independent of the possible number of future treated patients.

The summed benefit for treated patients in case the new drug is effective will depend not only on the weight one gives to the life-months of the patients and the estimated probability of success, but also on the number of treated patients. With an increasing number of patients the summed benefit—however it is measured—will increase linearly, if every patient has the same weight. So for the possible benefit one gets a series of straight lines with a certain slope, starting at the origin of the system.

The lines for the summed benefit for patients and for summed risk for subjects will have intersections. If the number of treated patients increases beyond this intersection point, the benefit is greater than the risk. In case the probabilities and the weights are equal for subjects and patients, one will have a gain if the same number of subjects is used in the trials. In case the slope of these lines is not so steep, the intersection point will be reached at higher numbers of patients. There is one important result: however small the gradient is, there must always be an intersection point, when the number of treated patients goes to infinity. This means that there will be always a gain, if there are many patients.

This figure indicates that one could try to develop a medical balance sheet along these lines for phase I trials, considering only the interests of the patients and of the subjects and not considering the profit or the scientific goals. Like the social budget of a company or of a nation, which has an increasing importance, such medical balance sheets could become a serious argument in the decision to start phase I trials. The weighing of the possible risks for approximately 50 subjects against the possible benefit of an effective treatment of patients for the next 20 years will nearly

always be in favor of the trial, especially if one takes the number of patients of future generations into account. I do not know what the weights should be in the individual case, but I know that the number of patients will increase in the next hundred years immensely. So we are still in the beginning of phase I trials in respect to the next hundred years. Generally speaking there should be many more trials, especially for frequent diseases with a bad prognosis—provided one weighs the future patients in the same way as the present subjects and one can convince the subjects to participate in trials for the benefit of future patients and generations—which could be developed into a strong ethical argument in favor of clinical trials.

In every case one should set up the goals of a phase I study initially as clearly and as simply as possible. One should always ask whether the goals are the right ones, and one should not try to solve all problems with one study.

2.2 Separate control system

The evaluation of evidence after each individual trial during phase I will be a decisive point for the system behavior. From the point of view of system analysis one should separate this function from the others. There should be an organization for performing the individual trials, defining the special goals, defining the trial, conducting it, and evaluating it. But it would be wise to separate the control system from this organization. The proposal is, to establish a definite organization and responsibility for the evaluation of evidence after each individual trial—which should not be identical with the persons responsible for performing the trials. The introduction of a separate control system for the whole phase I within a company would increase the overall performance.

Tasks of this control system could be:

—formulating and readjusting the stop- and go-conditions,
—evaluating the reports of individual trials,
—deciding on whether to proceed, to stop, or to start phase II.

2.3 How to structure phase I?

The formalization and the substructure of phase I could be improved. I have the impression that the complexity of organization and of structure varies considerably from company to company. The distinction between Ia and Ib trials might not be sufficient. Should there be only two or three phase I trials or about 8 to 12? One of the major problems is the question, how to structure phase I reasonably in order to get better performance. I do not know how this could be done, but it seems to be a problem which

should be mentioned and investigated. The degree of differentiation and complexity gives a hint for the empirically proven knowledge in an area. There is one optimal degree of complexity for a given situation and this degree of complexity could be pushed a little further for phase I.

2.4 How to measure performance of phase I?

This is to my knowledge an unsolved problem. One could try to evaluate a single trial, all trials within a definite phase I study for a certain drug, or a whole set of phase I systems. In case of the evaluation of a certain phase I study the following criteria are to be considered:

—the total cost, the time spent, and the number of subjects on the one hand,
—the impairment of subjects, the effectiveness of the new drug, the success on the market and the number of treated patients on the other hand.

How to combine such criteria remains an open question, but one should try to establish some formal criteria.

2.5 Selection of variables

The selection of measurements defines whether effects of the drug can be perceived or not. In case only a few variables are taken, there is no chance to detect drug effects on non-measured systems, as long as the effect does not immediately lead to a serious and obvious effect, which can be clinically perceived.

From a general point of view it is necessary to take the most sensible measurements and to try to check all physiological systems. It is wise to observe as many variables as possible when the substance is first applied to man, in order to maximize the information gathered with every subject. One must be alert to detect the unexpected during phase I, so a broad spectrum of variables should be sampled. This goal has its limits in the subject who is not willing to tolerate all kinds of examinations. Examinations which are not dangerous and painful and which can be performed easily should be used to a maximum extent at sensible time intervals. The more you measure, the more you might know and vice versa.

2.6 Intra-trial evaluation

The conditions for termination during an individual trial are sometimes not explicitly specified and the time between the treatment of a subject and the next subject might be not sufficient to evaluate all observations

properly. The intra-trial evaluation could be specified more precisely and the gathered information should be completely evaluated, before the next subject is taken—especially in the initial phases. Improvements in this respect could enhance a weak part of the present system.

2.7 Therapy in case of toxic effects

It seems to be rather seldom that toxic or dangerous side effects are perceived during phase I. However, in case of such effects an intensive care unit should be available. In preparing for such therapy there could be definite improvements, either by the specification of possible toxic effects, or by using special units as back-up, or by research in this area. In view of the fact that such cases are rather rare, one could question whether extensive precaution is reasonable. But one could also argue that this is not a question of money. The balance has to be found in every trial and general recommendations are missing. Providing the appropriate therapy in case of toxic or dangerous side effects could be one bottleneck for further development.

3. EXPERIMENTAL DESIGN

There is no specific theory available for experimental design during phase I. The design methods for phases II and III are used more or less sensibly. Since there are only a few subjects in phase I trials, these designs are not the optimal ones. I shall give two examples of experimental design, which seem to me useful for phase I. I then shall mention some problems, which could be further investigated.

3.1 Design for a single subject

When a drug is first applied to man, it is necessary to use some experimental design in this single subject. The situation of applying a drug for the first time in man is so unique that it is necessary to develop design models fitting to this situation, even if the design of experiments requires more than one experimental unit, requires random allocation of treatments to experimental units or some form of replication, which are not feasible in this situation.

Even if the usual instruments for experimental design are not applicable, there are some points of view which allow a design for the first single subject.

If we have only one single subject, we can not generalize to the population of all possible subjects, but we can generalize to the population of all

replications of the trial with this one subject. Our inferences are then restricted to this single subject, but they can be drawn using known statistical tests.

The control for the drug effect is the state before applying the drug. If we measure a variable on one subject before and after application of a single dose, we have the difference between the state before the drug and after the drug, but we have not yet a variance, with which this difference can be compared in order to make a statement whether the difference is significant at a certain level of probability.

The only way to get such a variance is to repeat measurements before and after the drug application. This repetition of measurements is a powerful instrument for the design with a single subject, nearly the only instrument we have. Unfortunately it is not often used. In the simplest case we have the following situation: A number of measurements are taken before and after the drug. The number of measurements need not be equal before and after the treatment, but we assume this here for simplicity. If we apply an analysis of variance model I to this situation, we can calculate the usual table for analysis of variance.

If the F-test is significant, this means that the difference between the measurements before and after treatment is larger than the variation between the replications. If for instance we measure the blood pressure ten times before and ten times after the treatment and the test is significant, this would mean that we have observed a change in blood pressure which is larger than our measuring error. We are not sure in this case that the difference can be attributed to the drug. We are only sure—at the specified level of significance—that the observed difference is larger than our measuring error, so we can maintain that we have observed a measurable difference. It remains a matter of judgment whether we attribute this to the drug. Such a conclusion is better than nothing and can be obtained with a single subject. If the difference is not significant, we can only say that the observed difference is of the same size order as our measuring error and an effect cannot be empirically shown.

Usually statisticians would not allow the application of analysis of variance to a single subject. The repeated measurements are correlated: there is no chance mechanism to allocate treatments, and the population for generalization is a theoretical one. However, this is the only way I know to apply known statistical techniques to a single subject. One has to redefine the assumptions: The difference between the measurements before and after treatment is assumed to be a fixed constant, the measurement-errors ε_{ij} are assumed to be distributed normally with mean zero and variance σ^2, and we have the well known model $x_{ij} = \mu + \eta_i + \varepsilon_{ij}$, where μ is the average value of the variable for this subject and η_i is the treatment effect. I cannot see why this model should not be applied. If the test is significant, we can

say that for *this* subject there is a measurable difference between before and after drug application.

By the same argument it is possible to use other and more refined analysis of variance models for a single subject, for instance using several measuring points in time (with orthogonal contrasts), using two error terms (for taking samples and for technical error), or using convariables or even MANOVA models. The inferences drawn are restricted to the single subject; however there are possible inferences, which is certainly better than nothing. The replication of measurements on certain points in time allows us with a single subject to use analysis of variance models.

3.2 Latin squares

If we have only a few subjects—more than one and less than 5 or 10—which is very often the case with phase I studies, the so-called Latin square is an appropriate design. Take for instance three subjects A, B, C, every one observed on three days I, II, III with three different dose levels 0, 1, 2. The doses are arranged in such a way that there is every dose in every row and every column only once. The three factors—subjects, days, and doses—are made orthogonal by this design. It is reasonable to assume that there are no interactions between days, subjects, and doses. An analysis of variance table can be calculated in the usual way, as indicated on the slide.

The restriction for these Latin squares is that the number of doses, the number of subjects, and the number of days must be equal, which can usually be achieved. This is a standard design for phase I trials.

3.3 Problems for further development

There is no general solution for *increasing dose levels* during phase I. If the estimated proper dose for the first application in man is 1, one could in a Latin square take the dose levels 0, ½, 1, and 2 which would require four subjects. If there is no effect, one might double the already used dose. But this raises the problem that one might miss the therapeutic dose range and find oneself, in one step, in the toxic area. If there is no effect, one should not always double the dose for the next trial. This depends on individual judgment. The strategy for increasing dose levels should be further investigated. Presently it is mainly decided by the individual experience of the investigator.

It is possible to combine the Latin square design with the idea of replication of measurements for the single cases. We have in this example 3 subjects on 3 days with 3 doses. On each of the occasions we measure before and after drug application and repeat the measurements 4 times. So we have $3 \times 3 \times 3 \times 2 \times 4 = 216$ observations. The simplest way of analysis

is then to use each dose separately, to perform an analysis of variance, and to test whether there is a treatment effect. With dose 0 there should be no effect. There are other models for analysis of such a data set, which can be omitted here. The combination of Latin square with repeated measurements before and after application of different drug doses seems to open a new and appropriate way for experimental design for phase I studies. This way should be developed further.

Even if there is today no adequate and systematic treatment of the possibilities for experimental design in phase I studies available—at least to my knowledge—there are some designs known, which can be applied to a single subject or to very few subjects. Further research in this area would be of vital importance.

4. STATISTICAL EVALUATION

The *statistical tests* should be applied in connection with the experimental design in the appropriate way. However, the test, whether there is a significant treatment effect or not, is not the only way of statistical evaluation. The statistical tests do not play the same role in phase I studies as they do in phase II and III studies for several reasons: The statistical model is confined to one or a few subjects, so the test means something different. We can test only one or very few hypotheses with a definite data set and the hypotheses have to be formulated before the design of the experiment. In phase I studies we are much more obliged to watch the unexpected, so we cannot formulate all hypotheses before hand and test them in the same trial. There must be ***hypothesis generation*** with the collected data to a maximum degree. So the methods of statistical hypothesis generation should be applied more extensively to phase I trials. One should describe the material as extensively as possible and observe every remarkable effect. One can, for instance, estimate the drug effects and give tolerance limits under definite side conditions in every subgroup of the collected data. One can use factor analysis of the differences between before and after drug application to derive hypotheses for the pattern of change induced by a certain drug. One could apply MANOVA design or other complicated models. The statistical squeezing out of the material seems to me to be allowed in the situation of phase I trials, provided it is well separated from the statistical tests and is described as such.

5. INTERACTIVE COMPUTER APPLICATIONS FOR PHASE I STUDIES

There is not enough space to develop the possibilities of computers for

phase I studies in detail. But I could try to give some ideas of for what purposes software systems could be developed.

5.1 Interactive statistical evaluation

The time between taking the measurement and getting the result in phase I studies could be considerably shortened by using computer systems in an interactive way. The data could be put into a computer by a terminal at the moment they are generated. The system could display the collected information, do any calculations, and give warning hints. To develop such a system one must specify the requirements for the statistical calculations, for the input and output. Especially for hypothesis generation such a dialog system could be a valuable instrument. One could incorporate simulations and put into the system estimates of variance from other trials. We have developed an interactive statistical evaluation system for the analysis of medical mass data, called SAVOD, and the experience with such a system is encouraging. A similar system could be developed for the special problems of interactive statistical evaluation for phase I studies. This could improve the results of phase I trials.

5.2 Control and guidance system

A different application would be to support the control of phase I by an interactive computer dialog. As shown earlier, the control of a phase I study should have its own organization. The formulating and the readjusting of the stop- and go-conditions could be improved, if a model for phase I is implemented according to the lines shown above. The results of different stop-and-go conditions could then be simulated, which would give further information for the guidance of phase I. It remains open whether such a system would be worthwhile. This depends mainly on the specification and on the requirements for such a system.

5.3 Simulation and model building

Since phase I puts heavy weight on hypothesis generation, one could try to develop a system for improving this task. The requirements for simulation and model building could be laid down. It is, for instance, conceivable to simulate a small subsystem of the human body and to simulate also the drug effect on this system with a computer program. Possible outcomes of further experiments could be predicted by such models and the predictions tested in real experiments. An ultimate goal would be a detailed simulation of the main physiologic subsystems of a subject participating in a phase I trial, doing this simulation with the individual parameters of

this subject and, in time, predicting the future behavior of details of the human subsystems while observing the real behavior. Such an on-line simulation of a subject during a phase I trial could increase the safety by showing dangerous effects earlier and could reduce the number of subjects necessary for phase I. The decreasing costs for computers and new technical and theoretical developments might make such applications possible.

The number of phase I studies will always be limited for obvious reasons. It is essential not to waste human and economic resources. In the long run intuition and individual experience are not the only guidelines for effective phase I trials.

Nature works by redundancy and by chance. This is also true for the fascinating process of developing new drugs, a process which is still in a very early phase of development. The sheer number of possible substances with some good effect together with the demand for new and effective treatment leads to ever new drugs.

We do not yet have a science in the strict sense to support these developments. We are collecting data, by trial and error, with individual experience and guidelines, like performing an art in conducting phase I trials.

Tomorrow there might be much more of a science in this field. At least a system theory for the development of a new drug, which deserves such a name, might be the goal. Some instruments are available today and this symposium might contribute to such a development.