Guidelines on the Statistical Analysis of Clinical Studies

CONTENTS

Chapter 1 Introduction.................................1

Chapter 2 Statistical analysis related items to be set forth in the clinical trial protocol

1. Objective of the trial................................. 2
2. Subjects................................................ 2
3. Group organization and allocation............... 3
4. Treatment............................................. 4
5. Observation and laboratory examination......... 6
6. Evaluation and criteria............................. 7
7. Discontinuance criteria............................. 8
8. Management of data quality....................... 9
9. Analysis method.................................... 9

Chapter 3 How to sort out data

1. How to sort out individual data.................. 12
2. Handling of problematic cases and incomplete cases.... 12
3. Multicenter cooperative trials.................... 13

Chapter 4 Analysis method and data management

1. Comparability of treatment groups................. 14
2. Description of statistical analysis............... 14
3. Incomplete cases and missing data............... 15
4. Multicenter cooperative trials.................... 16
5. Interim analysis data............................. 16
6. Incomplete termination of the trial.............. 16
7. Summary of results............................... 16

NOTES..................................................... 17

Chapter 1 Introduction
In drug effect evaluation, it is most important to make efforts to eliminate various biases at the stage of clinical practice, where necessary observation is made, and of data collection and analysis. It is therefore utmost essential to adopt a double-blind test with random allocation and appropriate statistical techniques for data analysis. Although this procedure is currently followed in drug effect evaluation, there still remain further improvements.

One of them is connected to a false judgment (evaluation) due to the improper use of statistical techniques. The present guidelines aim to provide the most suitable principles at this point of time by pointing out the misapplication of statistical technique, preventing it and thus assuring scientifically sound drug evaluation. Even though proper statistical techniques are used, valid drug effect evaluation cannot be achieved from data with poor quality and reliability. Some principles to be followed to improve the quality and reliability of data are already set forth in other guidelines (e.g., General Guidelines [draft], GCP). Furthermore, it is a principle to conduct comparative clinical studies at more than one institution (group), which secure the data applicability beyond the possible regional and inter-institutional variations.

The present guidelines deal mainly with parallel intergroup comparison in clinical studies designed for verification. For example, the crossover method indeed has theoretical values and, as a matter of fact, is considered the best way in clinical studies in patients with rare disease. In study conduct and evaluation, however, this method involves such difficult problems as the influence of a carry-over effect and changes in the baseline value in most clinical trials intended for drug effect evaluation.

Statistical inference can be roughly divided into two main categories: "test" and "estimation." Since statistical tests have been widely used, most explanation is given in this document by using them as examples. In general, however, usefulness estimation provides much information useful in clinical practice as to the degree of usefulness of the investigational drug rather than as to simply whether or not the investigational drug is more useful than the control drug. Therefore, it is desirable that point and interval estimation data be furnished where possible along with test data.

A new investigational drug should, in principle, be approved only when it is evaluated as being superior or equal to an already approved drug of the same kind and efficacy. It is statistically questionable to consider the investigational drug equal to the control drug on the ground that no significant difference is found between them at a comparative study. With consideration for clinical practicality, the use of the method for providing significant evidence to indicate that the investigational drug is not inferior to the control drug outside the allowable range is proposed in the present guidelines. Applicants are requested to utilize this method and further develop better way.

It is recommended that reliable drug effect evaluation documents of high quality be prepared in
Chapter 2 Statistical analysis related items to be set forth in the clinical trial protocol

1. Objective of clinical trials

The objective of clinical trials should be expressed clearly and concretely. Concrete expression must also be given to a secondary objective if any.

It should be noted that, when a clinical trial has too many objectives, statistical difficulties, such as the multiplicity of inference, may rise, resulting in a decreased amount of information on each individual objective [NOTE 1].

2. Subjects

1) Selection criteria

The intended subject population should be clearly defined as the conclusion obtained from the trial will be applied to this subject population. Since the definition of the clinical status of intended subjects is connected directly with the subject inclusion criteria, the status should be defined clearly in inclusion criteria (e.g., severity of disease, specific laboratory data).

The clinical status of subjects should be commonly specified concerning the following:

a) Normal healthy individuals or patients
b) Age range and sex
c) Admission status of patients (inpatients or outpatients)
d) Target disease, complication, and their severity (staging)
e) Medical history and presence or absence of complication and underlying disease
f) Details of preceding, prior and concomitant therapy

Description should also be given of the examination that is to be conducted to observe the above-mentioned conditions during the baseline period [NOTE 2].

2) Exclusion criteria

With consideration for safe conduct of a clinical trial, the subject exclusion criteria are established to
exclude, from the trial, subjects in which a bias in efficacy or safety evaluation because no reliable data could be obtained. In this instance, it is desirable to distinguish general rules from those peculiar to the investigational drug. Attention should be given to the safety peculiar to the investigational drug by referring to the precautions of the control drug if any.

3) Subjects to be excluded from trial

a) Subjects who can hardly be expected to respond to the investigational drug
b) Subjects in whom any main data are not obtainable or unreliable
c) Subjects who present symptoms easily indistinguishable from those of the target disease
d) Subjects whose condition makes it difficult to observe the rules for treatment and observation set forth in the protocol
e) Neonates, infants, children, and elderly subjects
f) Subjects with severe disorders of the liver, kidney, heart, lung, or the like (except the target disease)
g) Pregnant subjects, women of childbearing potential, and lactating mothers
h) Subjects with a history of drug allergy
i) Other subjects with a history of specific disease [NOTES 3 to 5]

3. Group organization and allocation

A summary of the protocol should be given. With the objective of the clinical trial in mind, the protocol must be properly formulated in terms of bias elimination, feasibility, and others.

1) Control group

Information should be given as to what type of subjects--untreated control, placebo control, positive drug control, or control of a different dosage level--has been chosen as controls simultaneously used.

In some dose-finding studies, a gradual dosage increase or decease regimen may be used under the clear rule for dosage increase or decrease; however, both of these regimens are detrimental when a symptom fluctuates greatly in severity even at the same dosage level owing to spontaneous variations in its course or when it is probable that such a rule will be violated.

Historical control is limited to special cases.

2) Group organization

Intergroup comparison, intra-subject laterality comparison, crossover comparison are involved in group organization. The method selected and justification should be specified.
3) Random allocation

In order to eliminate a systematic bias in study group allocation and assure comparable group formation on the basis of probability, randomization (random allocation) should be carried out. On this ground, statistical analysis is justified. The randomized design includes complete randomization, stratified randomization of important background factors (prognosis factors, co-variate), and a permutation block method which also assures the uniformity of short duration in process of time [NOTE 6].

In general, a double-blind randomization technique should be used in comparative studies. When no double-blind technique is applicable, however, an envelope method or method of allocation by telephone may be used. In this instance, the envelope method may easily have a bias in allocation (subject inclusion), the nature of treatment, and evaluation; therefore, extreme care must be used to eliminate bias.

4) Type of blind trials

The type of blind technique used (e.g., single-blind [subject], double-blind, or single-blind [data-analyzing investigator] technique) and specific procedure for its using (e.g., container labeling, double-dummy technique) should be described. This clearly reveals the intention of eliminating subject, observer, and/or investigator bias. Furthermore, description may also be given to the following: the condition and procedure necessary for breaking the code; and the code is broken for each individual subject or a portion or the whole of the trial.

If a recommended blind technique is not applicable or considered incomplete owing to apparent side effect occurrence, the reason and influence should be described.

5) Number of subjects

The target number of subjects should be mentioned, together with justifying statistical discussion or restrictions on study conduct. In the statistical discussion, the number of subjects or power of test should be clearly shown, along with the calculation procedure used, and the difference in efficacy between the drugs used in calculation and justification for it should be explained.

In the clinical trials intended to determine the clinical equivalence between the investigational drug and positive control drug, the clinically allowable difference (Δ) should be presented and the number of subjects used in the trial should be ascertained to have the power of test that permits the detection of differences greater than [NOTE 22].
6) Rule for trial completion

The rule for the completion of the whole trial (e.g., sequential analysis, number of subjects, duration of trial) should be described.

7) Inclusion period

The period during which subjects are selected on the basis of the subject inclusion criteria should be described.

8) Subject enrollment

Description should be given of the method for enrolling the subjects to be included in the clinical trial that for seizing the state of progress of clinical trial and state of adherence to the protocol, and that for ascertaining whether or not the subject inclusion criteria are fulfilled. In the randomized comparative trial where an allocation center is established, subjects in whom application for enrollment is made should be included in the trial when they fulfill the subject inclusion criteria.

4. Treatment

Time points of treatment initiation and treatment completion or discontinuance, as viewed as a whole, should be specified. According to the study schedule made out, the treatment and observation schedule for each individual subject should be described. As per this schedule, related information should be given concerning treatment for each period, such as details of treatment, dosage and administration, indications for dosage change (e.g., procedure for gradual increase or decrease in dosage), previous treatment, details of concomitant drugs/therapy (e.g., physiotherapy), and allowable limits. Such information should be determined at the stage of trial planning, with due consideration to the objective and feasibility.

1) Washout period

The necessity of washout, as viewed from the change in pathologic features due to the washout of previous treatment and efficacy evaluation in the absence of washout, should be described. Moreover, description should be given as to whether or not a placebo for washout is used and, if so, whether or not it differs in shape and color from the investigational drug. In trials where a washout period is established, the details of therapy during the period and reason for it should be described.

2) Baseline period
For purposes of the present guidelines, the baseline period is defined as the time interval from the instant the subject is considered to be included in the clinical trial to the instant investigational drug therapy is instituted. The significance of this period and observation and measurement to be made during this period should be described.

If it is necessary to standardize a method of therapy given during the baseline period, its details and reason for standardization should be described. In general, random allocation should follow the baseline period.

3) Duration of administration

The duration of administration should be determined with consideration for drug effect development and spontaneous changes in pathologic condition.

It is important to establish the duration required to determine a clinical effect, although this varies with the type of clinical trial, such as short-/long-term treatment and single/continuous/repeated dosing. Consideration should also be given to the resulting effects of single use, a baseline level change, drug tolerance occurrence, and other factors on clinical trial results. Concrete expression should be given by avoiding ambiguous expressions, such as "not less than one week or, if possible, not less than two weeks."

4) Dosage and administration

The dose, doses per day, dosage regimen, instructions for subjects, and other necessary information should be provided item by item in concrete terms. In the case of a variable dosage regimen, dosage change criteria should be specified.

5) Confirmation of compliance

The method and measures for the confirmation of compliance should be described. Possible compliance confirmation means include questioning, the use of a subject's diary, retrieval of unused drugs, and measurement of blood and urinary concentrations of a simultaneously dosed marker and drug ingredients. An effective method for avoiding noncompliance is to explain to the subjects about the treatment schedule in detail.

6) Follow-up period

The significance of a follow-up period, if applicable, should be described. The details of
standardization and reason for it should be described when a method of therapy given during the baseline period is to be standardized.

7) Concomitant drugs and therapy

(1) The requirements and allowable limits of concomitant drugs and therapy should be specified according to the objective of the study and nature of target disease, if necessary, for each of the baseline period, treatment period, and follow-up period, as follows.

a) Absolutely prohibited (e.g., drugs similar in structure and action to the investigational drug, drugs contraindicated);
b) Allowable at a certain dose for a certain period;
c) Prohibited for new institution
d) Discontinued after the subject comes to the study period;
e) Concomitant use allowed.

(2) Consideration should be given to the following if needed:

a) Prohibitive rules for the concomitant use of other drugs under development;
b) Rules for the treatment of underlying disease and single dose therapy for acute symptoms;
c) Rules for the treatment of side effects and complications;
d) Information on the use or dosage of a concomitant drug may be used as the index for drug effect evaluation in some trial designs. An example of this design is furnished by the dosing frequency of nitroglycerin in angina pectoris.

(3) Drug interactions

Whether or not the investigational drug may cause any interaction as in the case listed below, should be examined in determining the permissible limit of concomitant drugs. If any interaction is possible, precautionary notes should be made as precautions for drugs for previous therapy and concomitant drugs, together with those against side effects.

a) Pharmaceutical interactions: Turbidity of incompatible parenteral solutions, etc.
b) Pharmacological interactions: Reinforcement of blood coagulation inhibition by combined use of aspirin and an oral anticoagulant, etc.
c) Pharmacokinetic interactions: Absorption inhibition by combined use of a tetracycline and an antacid, etc.
d) Inhibitory actions on enzyme induction: Potentiation of tolbutamide by dicumarol, etc.
5. Observation and laboratory examination

With validity and reliability in mind, a description of the background factors to be recorded, observation/measurement parameters for efficacy or safety, and time points and methods of observation/measurement should be given. These time points should be clearly indicated. In this instance, time points may be specified by the elapse of time from the start of the trial and, where appropriate, in relation to the course of treatment.

In trials where more than one observation/measurement parameters for efficacy or safety are studied, how to handle these parameters should be discussed in advance.

1) Observation/measurement parameters for efficacy or safety

Observation/measurement parameters that are connected directly with the objective of the trial with reproducible results should be selected without excess or deficiency.

When efficacy and/or safety are evaluated in subjective categories or numerical scores, evaluations should be examined for validity. In this manner, efforts must be made so that satisfactory intra-observer reproducibility and inter-observer coincidence can be obtained.

2) Laboratory examination and physical examination

These examination items should be reduced to the minimum. If evaluation could be influenced by a difference in the method of measurement or representation among institutions, means to standardize data analysis, such as normal values in each individual institution, type of measuring instruments, observation and operation methods type of test reagents, measuring units, sample amount, and methods of description, should be stated clearly as needed. In trials where collective measurement is made at a specific institution, concrete description should be given of the method of specimen storage, package type, transportation method, test data delivery method, and the like, providing that the quality of the trial cannot be assured in the absence of such description.

3) Time points of observation and measurement

In the great majority of cases, observations and measurements should be made in time series; however, the time points of observation and measurement and allowable time span for each of them should be decided with consideration for the following and then specified.

a) Pattern of changes in signs, symptoms, and measured values: trend variation/recurring variation/random variation
b) Study design: fixed time point type/fixed response type  
c) Relation to administration time and pharmacokinetics  
d) Time intervals of observation and measurement in repeated observation  
e) Others (e.g., legal holidays, society meetings, annual events)  

4) Formulation and standardization  

The criteria for grading of changes in observation and measurement parameters and for presentation methods should be specified. When a measured value is recorded as it is, the significant figure should be stated in advance.  

5) Record form for medical personnel  

The record form for medical personnel may be used for recording compliance, side effects, and subjective and objective symptoms. It should be designed in a clear and practical format.  

6) Summary of the procedures  

A list showing the time points of observation and measurement and sequence of their importance should be made in advance. As for laboratory examination and physical examination, their operating procedures and methods should be summarized separately if necessary.  

6. Evaluation and criteria  

The time points and evaluation method (terminology and criteria) should be described clearly for individual evaluation items and general evaluation.  

1) Individual evaluation items  

The evaluation criteria should be specified for each evaluation item when evaluation is further made on the basis of observation and measurement data or laboratory and physical examination data. The time point of evaluation should be clearly described for the sequential evaluation items. Some evaluations are made subjectively by the physician-in-charge or subject and others mechanically on the basis of the prescribed criteria. How to handle each individual evaluation item in subjects who present few signs, symptoms, or laboratory or physical findings under consideration should be decided in advance, when appropriate.  

2) General evaluation
In coordinated consideration of more than one evaluation item, this evaluation should be made for each subject from the standpoints mentioned below:

a) Efficacy: Efficacy is assessed on the basis of the general improvement rating for changes in severity, course of pain alleviation, and other findings.
b) Safety: The side effects, abnormal laboratory data, remarkable laboratory changes, and causal relation to the drug are taken into consideration [NOTE 17].
c) Usefulness: Usefulness is assessed by the physician-in-charge in coordinated consideration of the efficacy and safety evaluations, QOL, and others.

In order to make these evaluations significant, they may be statistically examined for validity reproducibility, and coincidence in some cases. Observations made in time series should be analyzed or evaluated for some objectives [NOTE 13].

3) Assessment Committee Assessment Committee is organized as needed for the following purposes:

a) Criterion standardization and accuracy improvement of general evaluations;
b) Special knowledge is needed for side effect or other evaluation;
c) Rechecking is needed owing to the complexity of the evaluation criteria;
d) It is desirable to make collective evaluation, as in the case of evaluation by photographs and interpretation of recorded forms [NOTE 7];
e) Decision for discontinuance or continuance of the clinical study.

4) Evaluation of incomplete cases

The incomplete cases are classified according to the time of onset and the contents of an incomplete event. The event should be concretely examined for each subject. According to the objective of the trial, the subject in question should receive partial assessment [NOTE 16].

7. Discontinuance criteria

The requirements for subject withdrawal from treatment or observation, intervention after discontinuance, record keeping, and follow-up examination should be specified.

1) Requirements for discontinuance

In cases where the subjects are withdrawn from the clinical trial in the judgment of the physician-in-charge during the course of the trial, it is important not to overlook subjects to be included in efficacy, safety, or usefulness evaluation, taking the followings into account. Some cases of failure to return to the clinic may be considered the reason for discontinuance of the trial. The requirements for
discontinuance that can be set forth in the clinical trial protocol include the following:

a) Aggravation of symptoms;
b) The subject who is considered not to have responded to the test treatment;
c) Severe side effects or remarkable laboratory changes;
d) Aggravation of complications or occurrence of accidental symptoms (including unexpected accidents);
e) The change of treatment policy is needed;
f) Failure to observe the trial protocol for reasons other than described above;
g) The subject is found not to fulfill the subject inclusion criteria early in the course of the trial;

In addition, the subject cannot perform the clinical trial in the following cases:

h) The subject or his or her family member withdraws consent to participate in the trial;
i) Treatment is discontinued because the shifting of the physician-in-charge has made it impossible to transfer his or her business with exactness;
j) Other condition that indicates the cessation of treatment in the judgment of the physician-in-charge.

2) Record keeping

The reasons and time of discontinuance should be recorded. In subjects withdrawn from the trial owing to the occurrence of a side effect or abnormal laboratory data, clear description should be given so that the details and results of follow-up examination after discontinuance will be kept as a record.

8. Management of data quality [NOTE 8]

The management of data quality is of importance for the assurance of reliability in clinical trials. Missing values, outliers delay in the time of evaluation, and state of compliance are involved in the quality evaluation of clinical trials. In inputting these data, it is important that the input method be specified beforehand in concrete terms. Comparative studies are based on the premise that random allocation is carried out with exactness. Moreover, attention must be paid to the confirmation of consistency among evaluation items and the like. Efforts must be made to assure the reproducibility and reliability of data obtained.

9. Analysis method

In clinical studies, since there is a possibility that incomplete cases occur, complete data are not always obtained. These incomplete cases cause bias in data analysis; therefore, a guideline for handling of incomplete cases in classification by reason and analysis should be established beforehand according
to the objective of the clinical trial.

How to cope with a background factor-related imbalances that make it impossible to make comparison among treated groups should be examined beforehand [NOTE 20].

The main statistical analysis used to verify a drug effect should be clearly described in distinction from secondary statistical analyses. In general, the main statistical analysis should be made for each of the efficacy, safety, and usefulness evaluations. As for the main statistical analysis, the statistical hypothesis and analytical technique suitable for its verification should be selected beforehand and described, together with the analysis items. In this instance, consideration should be given to the following: the definition of the rate of safety, efficacy and usefulness (meanings of a numerator and denominator) together with definition of subgroup; whether or not subgroups are analyzed separately; what points of time are taken into account; to which comparison much importance is attached in the case of more than two treated groups. When more than one statistical inference is drawn as the main statistical analyses, attention must be focused to the level of significance and confidence coefficient [NOTES 9 to 14]. In clinical trials whose objective is to verify the clinical equivalence to the positive control drug, clear description should be given of the analysis method that presents evidence to show that the investigational drug is not inferior in efficacy to the positive control drug by more than maximum clinically allowable difference (Δ) [NOTE 22].

1) Coping with incomplete cases

Handling of incomplete cases may cause a great bias to study results or badly affect the reliability of data. With these points in mind, the incomplete cases should be classified according to reason and, at the same time, a guideline for handling in analysis should be described clearly according to the nature of the trial--that is, investigative versus practical or exploratory versus verificatory [NOTES 15 and 16].

2) Coping with background factors

The subject background factors to be recorded are known to be related clinically to prognosis; therefore, the imbalance in background factors between treated groups may influence the results of a comparative study. Prior to statistical analysis of drug effects, whether or not background factors recorded are comparable between treated groups should be examined and the comparison procedure should be specified in advance.

In statistical analysis of background factors, it is of importance that examination is conducted not by evaluation on the basis of significance levels of 5% and 1% or confidence coefficient levels of 95% and 99%, but from the viewpoint of the influence of the imbalance in background factors on efficacy,
safety, and usefulness evaluations. In great majority of cases, careful attention must also be paid to the imbalance in background factors in about 15% probability of significance. Furthermore, examination of imbalances of different kinds by more than one statistical technique is also recommended.

How to cope with imbalances in background factors must be examined beforehand. When statistical correction may be made, the analysis technique should be examined beforehand.

3) Analysis items

In general, the main statistical analysis items are individual evaluations of efficacy, safety, and usefulness and their general evaluation. Of the observation/measurement parameters and laboratory examination/physical test and evaluation items, the analysis items tailored to the objective of the clinical trial may also be selected beforehand. With respect to handling these data, however, it is necessary to make a thorough examination.

Observation/measurement parameters and laboratory/physical test and evaluation items except the main analysis items should be analyzed thoroughly as secondary statistical analyses. These statistical analyses may not only support the main statistical analyses, but also suggest a new clinical hypothesis for the investigational drug. It is desirable to examine a secondary statistical analysis in advance, for example, by planning it in proper quantities and containing it in the clinical trial protocol.

4) Level of significance and confidence coefficient

The level of significance, confidence coefficient, and two-tailed or one-tailed test during analysis should be specified. Usually, 5% and 1% are used as the levels of significance and 95% and 99% as confidence coefficient. In testing, it is recommendable to describe not only whether or not the difference is significant, but also the significant probability (p value).

5) Analysis technique

The analysis technique must be selected in accordance with the study design and evaluation scale. For example, when many kinds of tests are repeated without any fixed principle in the two-group comparison of serially classified data, probability of errors of the first kind is largely increased. In multiple group method, the multiple comparison tailored to the objective is more rational than an intergroup uniformity test. In a dose group comparison, use may be made of the multiple comparison where consideration is given to the order of level and analysis of dose-response relationship.
6) Person in charge of statistical analysis

The name of the investigator in charge of statistical analysis, institute, and where to contact should be described clearly.

Chapter 3 How to sort out data

It is necessary to sort out the collected data, including not only those from the subjects included in analysis but also those from the subjects excluded from analysis.

1. How to sort out individual data

Randomized subjects may not be excluded from the data list even when they are not under study during the study period during which the investigational drug or placebo is administered.

Subjects who are not randomized because of failure in conforming to selection criteria need not to be placed on the data list. A unique identification number for each subject should be used in the representation of all data. The following data must be furnished for individual subjects:

1) Demographic data

Age, sex, height, body weight, etc.

2) Important background factors

Severity, laboratory data, complications, effect of prior treatment, etc.

3) Dosage of the investigational drug and duration of treatment days (when the dosage is increased or decreased according to the drug effect, each dosage level, final dosage level, and number of treatment days)

4) Dates of trial initiation and completion

5) Compliance

6) Evaluation items and main time points of evaluation concerning efficacy, safety, and usefulness
7) Whether or not the subject in question is included in efficacy, safety, and usefulness evaluation

8) Concomitant symptoms (including abnormal laboratory data and abnormal changes)

   a) Nature of symptom
   b) Time of onset
   c) Relationship to the investigational drug (e.g., definite, probable, possible, or unrelated)
   d) Severity (e.g., severe, moderate, mild)
   e) Outcome: death, cessation of treatment with the investigational drug, change in dosage, treatment continuance without modifying dosage, etc.
   f) Dosage at onset
   g) Cumulative dosage until onset, especially in cases where it is so severe that treatment need to be discontinue
   h) In cases where more than one side effect is elicited in one subject, an outline of the subject in question should be given to show the nature and onset of side effects [see Chapter 4 Section 2.(5)].

2. Handling of problematic cases and incomplete cases

   A list of all problematic cases and incomplete cases (ineligible, discontinuance, dropout, failure of follow-up, etc.) in which the items described in sections (1) to (6) below are contained should be made.

   1) Details of problem;
   2) Time of outset of problem;
   3) Dosage at the outset of problem;
   4) Concomitant drugs at the outset of problem;
   5) Reason for the problem outset;
   6) Coping with incomplete cases [NOTES 18 and 19].

3. Multicenter cooperative trials

   Efficacy, safety, and usefulness data should be sorted out for each physician-in-charge. Efficacy, safety, and usefulness evaluations should be sorted out on a physician-basis, particularly when physicians have charge of many subjects.

Chapter 4 Analysis method and data management
The analysis result should contain the consideration of the comparability of treated groups and statistical analysis including the handling of problematic cases and incomplete cases. In the case of multicenter cooperative trials, a discussion about the consistency of evaluations should also be contained.

1. Comparability of treatment groups

1) The comparability between treated groups should be shown in tables, graphs, or other statistical methods with respect to the background factors of subjects, number of subjects included in analysis, and number of subjects who successfully complete the clinical trial. The background factors to be considered vary with the characteristic of the target disease. Commonly, they included demographic factors (e.g., age, sex), duration of illness, severity of disease, baseline levels, previous treatment.

2) When a background factor differs from one treated group to another, the influence of the fact on further data analysis and data interpretation should be described. If important background factors are statistically adjusted, the reason for the selection of them should be explained [NOTE 20].

3) When conventional controls are used, explanation should be given of the following: in what manner the control group has been selected; what literature search for other candidates has been made, and whether or not the conclusion obtained from them is similar to that from the controls used praktically.

2. Description of statistical analysis

1) Results of the main statistical analysis

Analysis results on efficacy, safety, and usefulness should be represented with the information listed below:

a) Analysis items
b) Analysis technique
c) Statistical inference and test results
d) Estimated results of the main characteristic data (e.g., mean, incidence)

In particular, items b), c) and d) above should be described in statistical terms and consideration should be given to the power of the test, estimation accuracy, and the like.

2) Results of secondary statistical analysis

Data of secondary statistical analysis that support those of verificatory statistical analysis should be
represented in proper quantities.

Furthermore, the eminent facts found by exploratory statistical analysis should be indicated. These facts may present a clinical hypothesis to be verified by a new clinical trial.

3) Analysis procedure

a) Course of computation

Summarized documents should be furnished as needed so as to permit rapid and easy confirmation of the course of analysis.

b) Test

Data on test statistic, significance level, and significant probability (p value) should be sorted out. Clear description should be given as to whether the level of significance and p value are one-tailed or two-tailed [NOTE 21].

c) Estimation

The point estimation value, confidence interval, and their confidence coefficient values should be indicated.

d) Figures and tables

Use of appropriate figures and tables makes the presentation of the data more intelligible. In such figures and tables, either standard deviation or standard error should be described as needed.

4) Clinical equivalence

In trials where a control agent (positive control drug) except placebo is used, the efficacy and usefulness of the investigational drug may be alleged on the ground of the similarity noted between the two drugs. In this instance, active evidence must be provided to indicate that the investigational drug is equal or superior in efficacy and usefulness to the positive control drug [NOTE 22]. In some cases, the following must be given in addition to the comparison with the positive control drug in the clinical trial in question: the effect level and consistency of the positive control drug in similar studies that have been conducted; or a ground indicating that the efficacy of the investigational drug has been judged stronger than the minimum efficacy necessary for the placebo used under similar study conditions.
5) Side effects

The following should be given concerning side effects for each treated group:

a) Nature of side effects, number of events for each side effect, number of subjects with side effects, and incidence;
b) Comparison by statistical analysis of the incidences of side effects of treated groups. Correlation of side effects to subject's characteristics, such as body weight, age, sex, concomitant drug, dosage of the investigational drug, duration of administration, abnormal laboratory data, or abnormal changes, should be determined as needed.
c) The same rule applies to abnormal laboratory data and abnormal changes in laboratory data.

3. Incomplete cases and missing data

In general, incomplete cases and missing data constitute potential factors of bias in clinical trials; therefore, to what degree handling of incomplete cases and missing data influences trial results should be considered. Commonly, statistical analysis of the subjects who are entered in the trial or included in efficacy evaluation should be made according to the objective of study.

4. Multicenter cooperative trials

When the results of a multicenter cooperative trial are to be marshaled collectively, the validity of the procedure should be confirmed. In other words, whether or not results are highly consistent even in different physicians-in-charge should be considered. Finding evaluations with large variations, between institutions or that the evaluation by a physician is extremely different from that by others should be noted and discussed.

5. Interim analysis data

When statistical analysis is made during the course of a clinical trial, the data obtained should be furnished to describe their effect on the progress of the clinical trial and final conclusion. If interim analysis has been made in spite of no description in the protocol, the reason for the analysis should be explained.

6. Incomplete termination of the trial

If it is decided to close down the trial prior to achieving the target number of subjects, reasons for this should be documented and the implementations to the results discussed.
7. Summary of results

The clinical characteristics should be summarized from the statistical analysis results obtained.

NOTES

1. Multiplicity of inference

There is ample room for selection as to what analysis method is used for sufficient drug action data obtained from clinical trials. At present, many statistical inferences are drawn for a large number of data obtained from a clinical trial. When many statistical inferences are drawn, however, the probability that a significant result is yielded is apparently increased, resulting in an increased probability of errors of the first kind. This problem is discussed as multiplicity in clinical trials.

The principal matters of multiplicity discussed at present with respect to comparative trials in Japan are as follows:

a) Multiple group (drug) comparison

Multiple group comparison of more than 2 groups (drugs) is made, for example, by pairing 2 groups (drugs) each see NOTE 9.

b) Rounding of serially classified data

The contingency table is marked off at various points and tests are then repeated [see NOTE 3].

c) Multiple tests

Several different tests are conducted on a single set of data [see NOTE 11].

d) Multinominal comparison

Comparison is made individually for the observation/measurement parameters and laboratory/physical test and evaluation items [see NOTE 12].

e) Time-course comparison
Comparison is made at each time point of evaluation to determine whether or not there is a time-course disparity in drug effects [see NOTE 13].

f) Subgroup analysis

Comparison is made in many strata, such as sex, age, disease type, and severity [see NOTE 14].

g) Interim analysis

Analysis is made at many time points during the course of the trial.

2. Eligibility

It is helpful to ascertain whether or not a subject fulfill the subject inclusion criteria by the use of an eligibility check sheet prior to the entry into the trial.

3. Complications

Such complications as disorders of the liver, kidney, heart, and lung depend on whether or not the disease is identified and whether or not treatment is indicated. They may also be specified with the values for the main examination items.

4. Pregnant women and lactating mothers

The propriety of administration to pregnant women and lactating mothers is judged based on study data on reproduction, transfer from mother to fetus, and excretion in milk. In the case of specific drugs, however, it may be necessary to arrange contraception.

5. Subjects

In some instances, it may be necessary to describe environmental factors (geographical, social) and cooperation of subjects (or their guardians) (e.g., cases of difficulty in returning to the clinic after trial initiation, as in the case of subjects coming a long way during the longterm follow-up).

6. Stratified random allocation

In the case of the factors supposed to greatly influence prognosis, the stratum balance can be adjusted by stratified random allocation. Excessive stratification, however, may cause estimation efficiency to decrease; therefore, the necessity and rationality of stratification should be demonstrated beforehand.
In multicenter clinical trials, it is common to prepare a permutation block for each institution with institutions as a stratum. A minimization method and the like are proposed to adjust the balance with respect to more than one important background factor, but special care must be exercised in analysis.

7. Evaluation by the Committee

Evaluation by the Committee is uniformly made on all data at the same time, therefore, the evaluation improved in reproducibility and consistency, but the data peculiar to an individual subject which can only be evaluated by the treating physician may be ignored.

8. Management of data quality

The quality of a clinical trial depends on the quality of data. When a clinical trial is conducted in accordance with the protocol and followed by accurate inquiry, observation/measurement, and evaluation, and correct analysis and interpretation, a valid and reliable conclusion is obtained. When taken in a narrow sense, data quality control means that errors in data collection are detected and corrected. This procedure can be divided into several stages.

The case records and related material should be retrieved immediately following a subject's completion of the trial, although this requirement varies with the scale and period of the clinical trial. In some cases, they may be retrieved in divided portions, such as a registration record, a diagnostic examination record, a treatment record, and a follow-up record. This makes it easy to manage the quality of data.

Most case records are coded, inputted to a computer, compiled, and then summarized by statistical analysis. At the stage of case recording, data should be checked first for missing and delay in the prescribed time of inquiry, observation, measurement, and evaluation, and special codes are assigned to them if necessary. The input data are then read out in collation. Although checking can also be done by examining whether or not a remainder of zero is obtained by the subtraction of data independently inputted, cleaning at the stage of case recording or coding is of significance because erroneous input of handwritten figures may occur repeatedly.

The input data must be closely examined according to the nature of data. Ranges are checked as follows: for example, when the specified age range is from 20 to under 65 years, subjects whose ages are outside this range are picked out. Laboratory data are checked similarly and the outside-the-range values are picked out. In logic checking, for example, subjects in whom the systolic blood pressure less diastolic blood pressure is negative or subjects with no side effect in whom the trial is reported to have been discontinued owing to side effect occurrence are picked out. When an improvement rating is determined collectively from several observation and measurement evaluations,
it may also be checked for validity. Graphic representation is useful for editorial checking and
abnormalities are often found by plotting against time.

The data which are picked up by those checks are not necessarily all wrong. In many cases, errors
can be corrected by referring back to the case records and where necessary the investigator. The
longer the elapsed time from entry to checking the more difficult verification of data becomes. As it
cannot be guaranteed that all data can be corrected by this extra checking, it is better to be careful at
the time of initial data collections particularly for principal data.

9. Multiple group comparison

In most analyses of clinical trials using more than two groups, comparison has been repeated between
the pairs, as in the case of two-group comparative studies. In this instance, the level of significance in
statistical tests is $\alpha$ in each individual pair comparison, but the overall level of significance for the
whole consisting of plural pairs is a value much larger than the established value of $\alpha$. In
verificatory data analysis, in order to give assurance that a statistic is statistically significant at 5%
level, a correct point of significance must be selected so that the overall significance level of 5% is
obtained.

In this multiple group comparison, the use of uniformity tests (e.g., F-test, Kruskal-Wallis test) or
multiple comparison is involved. Uniformity tests, however, are used to examine whether or not the
effects of plural groups are equal to one another. In addition, this method is not intended to compare
a specific pair of clinical concern. For this reason, even when the null hypothesis that "the effects of
a pair is equal to each other" is denied by a uniformity test, where a difference is found among
multiple groups is uncertain. Furthermore, the possibility that a difference between a pair of concern
is overlooked is likely. It is desirable to make multiple comparison according to the hypothetical
structure of a clinical trial.

These multiple comparison methods include Dunnett's method (suitable for comparing a standard
group with all other groups), Tukey's method (used when there is any concern in all pair comparisons),
and Scheffe's method (a more general method). Moreover, a conservative approximation method
(using Bonferroni's inequality) can be employed when no accurate point of significance is obtained.
These methods are used not only for continuous data, but also categorical data. The interval
estimation corresponding to multiple comparison is designated as joint estimation.

Multiple group comparison, even when made by a suitable method, may pose a statistically
complicated problem, such as a wrong interpretation of results and lessened power of test. In phase
III clinical trials for a verificatory purpose, therefore, it is desirable to make a two-group comparison
with a positive control or placebo where possible. When comparison is to be made among more than
two groups with a sufficient reason, the justification of establishing the number of subjects and analysis plan must be described in the clinical trial protocol.

10. Rounding of serially classified data

For example, after a number of quadrant contingency tables are compiled by partitioning the final general improvement rating into four categories ("markedly improved," "moderately improved or better," "slightly improved or better." and "aggravated or worse") the routine $\chi^2$-test may be repeated. Clinically, information on the proportion of each category to total is understandable information of importance; however, such repeated testing may cause the probability of producing errors of the first kind to increase. In this instance, for example, the use of maximum $\chi^2$-method allows the probability of significance to be exactly computed.

11. Multiple tests

Tests of more than one kind, such as Mann-Whitney U-test and $\chi^2$-test, may be applied to the final general improvement rating and contingency table for usefulness. In this instance, it is also natural to increase the probability of producing errors of the first kind. Use of Wilcoxon two-sample test (Mann-Whitney U-test), cumulative $\chi^2$-method, and maximum $\chi^2$-method in serially classified data is proposed, and features of each method are elucidated. For instance, U-test shows the high power of test for average differences in drug effects. The maximum $\chi^2$-method is suitable when interest lies in drug effect differences at the clinically important step ("aggravated" or "markedly improved") on the assessment scale. The cumulative $\chi^2$-method possesses the temperate power of test for the degree and distribution type of efficacy variations as well as average drug effect differences.

When these different tests are used in comparative clinical trial data, results of statistical inference are known to show a difference not infrequently.

12. Multinominal tests

In the case of a very large number of comparisons concerning more than one observation/measurement parameter, laboratory/physical test item, or evaluation item, the overall probability of producing errors of the first kind increases even when individual tests are kept at the level of significance. This increase varies with the correlation structure between the items under consideration. For example, when two drugs are equal in 20 items to each other and no correlation is found between any two items, there is a significant difference in 64% probability. In contrast, the probability shows about 22% when correlation coefficient values between any two items are 0.8. In the case of very good correlation between items, an increase in errors of the first kind is slight, but the increase becomes a
momentous problem when a drug effect difference is to be verified in many items with a low degree of correlation.

It is difficult to infer correlation structure among items in advance. This makes it difficult to control errors of the first kind in multinominal tests.

One way of coping with this problem is to depress errors of the first kind by correcting the point or probability of significance in individual analysis testing. In this connection, the correction after Bonferroni in which an uncorrelated state is assumed may actually become an excessively conservative method. In general analysis, on the other hand, one way is to avoid difficulty in statistical adjustment by predetermining how to cope with plural measurements collectively. In this instance, however, it is important to examine whether or not the collective method is medically appropriate. Furthermore, a test method for the plural test statistics summarized in a small number of data by figuring out the weighted sum of the statistics is also proposed.

In phase III clinical trials intended to verify the effect of an investigational drug, it is necessary to choose a small number of medically important items out of many beforehand and specify them in the protocol. When comparison is required in plural items probably correlated in a low degree, a proper statistical adjustment must be made on the basis of the rules for results from plural items, laid down beforehand.

On the other hand, when data are analyzed not for the verification of drug effect differences but for the exploration after a comparative study intended to find the nature of effects to be verified in future, it is important to build up a proper hypothesis with consideration for the number of items with a significant difference and level of significant probability.

13. Time-course comparison

In comparative studies, the subjects receive assessment and measurement of every kind at various times to evaluate drug effect differences. In this instance, each drug group is liable to repetitively undergo a test used when data at the time of treatment initiation and at subsequent time points of evaluation correspond to each other or when the two groups do not correspond. However, this causes the probability of producing errors of the first kind to increase. In this case, a special method for comparing a trend or a profile, as opposed to point of time data, is necessary.

14. Subgroup analysis

In some cases, analysis is intended to verify the effect of an investigational drug in a certain stratum (subgroup) of background factor. The question of repetition-induced multiplicity as a result of
statistical inference for each individual subgroup is one of the most difficult problems. Concerning the handling of correlation structure among background factors, an unsolved problem is posed, as in the case of multinominal comparison, when subgroup analysis is repeated for each of the plural background factors, an increase in the probability of producing errors of the second kind, expressed as a decrease in the power of test due to a decreased number of subjects, also becomes a problem. In subgroup analysis with a verificatory purpose, it is necessary to secure the sufficient number of subjects for each group from the stage of planning.

Information as to interaction, expressed by a subgroup (e.g., sex, age)-dependent difference in the mode of effects between investigational and control drugs among subgroups, is supremely important for the evaluation of the effect of an investigational drug. In trials by comparison between an investigational drug and a positive control drug similar in the type of action mechanism to it, the interaction is difficult to presume in the great majority of cases unless a thorough examination is made on results of nonclinical studies and previous clinical studies of the investigational drug. In the case of verificatory data analysis, it is essential to select out a few background factors from among many for the purpose of stratification in advance. In this instance, previous data on the investigational drug should be considered together, and attention should be focused on the interaction between the drug and background factors.

15. Classification of incomplete cases

For example, the incomplete cases should be roughly divided into two categories--one being ineligible, related to a predrug problem, and the other being discontinuance, dropout, and deviation (violation), related to treatment after trial initiation.

a) Ineligible

The subjects who no not fulfill the subject selection criteria or who have any of the conditions of the subject exclusion criteria should be considered ineligible for trial entry. When found to have nontarget disease or not to fulfill the subject selection criteria, subjects are divided into two categories one being related to the target disease (e.g., incomplete data on diagnostic evidence), and the other being related to the background characteristic and factor of the subject (e.g., age, history, complication, and baseline data).

b) Discontinuance

This is the case where in the medical judgment of the physician-in-charge, treatment of a subject is discontinued deliberately. This case does not always imply protocol violation, being considered a complete case if circumstances require. As a matter of fact, this category includes cases where
treatment is changed or terminated owing to any side effect, aggravation, cure, or accidental symptom and where the subject is found ineligible after trial initiation. In some cases, dosage regimen is changed or concomitant therapy is given, with the result that treatment is discontinued. If treatment is discontinued owing to side effect occurrence or cure, the details of this event should be recorded exactly at the time of discontinuance. The resulting record provides important information because the event indicates the important outcome of the study treatment.

e) Dropout (failure of follow-up)

This category is formed by cases where circumstances unrelated directly to the trial treatment (e.g., subject's moving, pressure of business, uncooperative attitude toward the trial) make it impossible to carry out observation/measurement, continual examination, and evaluation, thereby cutting off data. On the other hand, data may also be cut off for reasons related to study treatment (e.g., a risk of side effects, alleviation of the target disease or symptoms) via noncompliance. In this instance, if the judgment of the subject is ascertained to be medically suitable, this case may be considered discontinuance (i.e., study treatment is terminated). Although dropouts may be used synonymously with failures of follow-up, it may be proper that the dropouts in whom no information is obtained during the follow-up period be the latter. A combination of discontinuance and dropout may be designated as incomplete cases. These subjects are characterized by the cutoff of data.

d) Treatment violation

A subject is selected and then entered into the trial. In the case of comparative trials, the study treatment is instituted according to the specified allocation. After treatment initiation, concomitant drugs, a deviation from the clinical trial protocol may occur with respect to concomitant drugs, concomitant therapy, methods or time points of observation/measurement, examination, or evaluation. These cases are considered treatment violation. This includes the modification of dosage regimen and addition of a prohibited concomitant drug. The violation may also be caused by inadequate management of the clinical trial and overlapping between the scheduled day of examination on one hand and a holiday and the day of association meeting on the other.

e) Noncompliance

Cases where a subject does not take the medicine under subject compliance instructions or withdraws from the trial are considered noncompliance. In most cases, delay in returning to the clinic causes the subject to take the medicine, so that both observation/measurement schedule and medication are violated.

Noncompliance is caused by the following cases: the subject fails merely to take the medication; he or
she has a fear of side effects; and, in his or her judgment, the medication is ineffective or the disease under consideration has completely cured. A probable measure to counter noncompliance is to exclude uncooperative subjects in advance by establishing an observation/measurement period and a trial period. This method, however, may cause the subject group members to change.

In the case of subject's concern about side effects, a follow-up examination reveals that the subject is concerned too much about side effects or that, from a clinical viewpoint, the concern is derived from an apparent side effect.

Violations on the part of the physician and noncompliance on the part of the patient do not necessarily result in cessation of the trial although some data are missing or inappropriate. If any violation persists beyond a certain point of time, however, it becomes synonymous with cessation.

16. Handing of incomplete cases at the stage of analysis

Problems related to incomplete cases in comparative trials should be tackled from various viewpoints: whether the problem has occurred early or late in the course of the trial; whether or not the cause is related directly to the trial treatment; whether the reason lies in the part of the physician-in-charge or part of the subject: and whether the problem is involved with the whole or part of the trial. Furthermore, the handling of subjects may vary with the objective, period, and scale of the trial, and state of blindness (double-blind manner or not). In handling the data from incomplete cases, the most careful attention should be given to the following two: (1) bias interference with trial results (purpose of random allocation, intended to maintain the impartiality of trial, may be defeated when the bias is strong); and (2) distortion of the subject population assumed at the stage of planning. It is not enough to simply count the number of ineligibles, subject withdrawals, and dropouts and to compare the treated and control group in relation to their respective percentages to the total. In addition to this procedure, it is necessary to make more detailed comments as well as to prepare the material giving a detailed account of how individual incomplete cases have been handled.

Naturally, the decision as to how strictly the trial plan is specified depends on the objective and nature of the trial. The decision is also related closely to the judgement as to whether or not data from incomplete cases are adopted. The objective and nature of the trial are classified, for example, as exploratory or verificatory, or investigatory or practical, but the valid handling of data varies depending on their respective standpoints.

From the very nature of the case, the guidelines for handling incomplete cases should be determined not on a case-by-case basis after completion of the trial, but in advance. The number of incomplete cases is an index suggestive of the quality level of a trial, to say the least; therefore, large numbers mean that the reliability of the trial is lost.
a) Investigatory trials and practical trials

Investigatory trials should be conducted not only in clinical practice but also in the field where true homogeneity is obtained. Based on this recognition, only subjects eligible for entry into the trial who have completed the present trial in accordance with the protocol should be strictly selected and used as the numerator and denominator, for example, when the rate of improvement is taken into account (protocol-compatible).

In the case of practical trials, on the other hand it is also claimed that the subjects who have undergone the trial treatment should be included in analysis regardless of duration of treatment—that is, without considering incomplete cases to be violators by rote—and added to the denominator even if there is no possibility that they will show up in the numerator as the subjects evaluated as being "improved" (intent-to-treat).

These two extremes of opinion differ apparently from each other in the definition and objective of the assumed subject population. Emphasis is put on the comparison of treatment regimens and investigation and explanation of pharmacological features under certain limited conditions in investigatory trials, whereas practical trials give priority to the decision as to what treatment is practically selected out from among the trial treatments given in a comparative trial. Presumably, usual comparative clinical trials are of a nature intermediate between these two kinds of trials with high frequency. In short, after the reason for incompleteness is inquired at the time of its occurrence, it is more realistic, for example, to exclude a subject (or included as an "unassessable" subject) in efficacy evaluation but include the subject in safety and usefulness evaluation according to the type of data. Nevertheless, this method is inappropriate, for example, to subjects in whom treatment of too short duration for some reason other than side effect occurrence makes it impossible to make efficacy evaluation, because there is a risk that safety rating may be rated higher than actual.

b) Details of ineligibles

Identification of an ineligible largely depends upon whether or not the trial has been conducted in a double-blind manner. In a non-blind trial using, for example, the sealed envelope method, study treatment is not given under the specified instructions in some cases. These cases are defined as treatment violation. Once the reason for violation has been learned after the trial, it is generally difficult to make comparative evaluation so as to eliminate bias. Although it is, of course, unreasonable to handle a slight deviation in age restriction and mistaken diagnosis indiscriminately, not a few ineligibles, even when due to a mistaken diagnosis, may be included in safety evaluation in spite of rejection from efficacy and usefulness evaluation.
Subjects with disorders of the kidney or liver may have side effects; therefore, these conditions are set forth in the subject exclusion criteria. Nevertheless, if such subjects are enrolled in the trial, special handling must be done in relation to the actual occurrence of side effects.

c) Reasons for discontinuance and dropout

Some data, to say the least, can be included in analysis if such a reason for discontinuance as side effect occurrence or cure is clear. In subjects in whom a trial is discontinued owing to side effect occurrence early in the course of the trial, usefulness as well as safety may be evaluated although efficacy evaluation is considered unreasonable in view of the investigatory nature of trial. In Europe and the United States where the concept of usefulness is minor, some investigators are of the opinion that such subjects should be evaluated as being ineffective and included in efficacy analysis, from the viewpoint of a practical trial. Other opinions hold that, in computing the rate of improvement, the subjects of this kind should be added to the denominator in the sense of a penalty for side effect occurrence, thereby lowering the rate.

In a trial where the subjects are treated for four weeks and then followed up for one year, it is certainly unreasonable to make efficacy evaluation on subjects withdrawn from treatment at 2 weeks. Nevertheless, such subjects do not appear on the surface when considered simply ineligible or incomplete; therefore, it is appropriate to include them in usefulness evaluation.

When fracture due to a fall during the treatment period makes it impossible for subjects to visit the hospital so that they are handled as dropouts, some of them are not considered simple dropouts because the drug used may have a central nervous system depressive action or influence bone metabolism. The same is true in the case of discontinuance due to a traffic accident.

d) Deviation or missing measurements

In investigatory trials, data on treatment violations and noncompliance should, in principle, be excluded in analysis and an undue delay in observation/measurement examination, or evaluation, as viewed from a common-sense standpoint, become a problem. In practical trials, on the other hand, the data included in analysis cover a considerably wider range. In either case, acceptance or rejection of data is decided under a fixed arrangement. It must be noted that, in this instance, there is a constant risk that bias may interfere with trial results and that this bias cannot be simply evaluated.

Generally speaking, subject exclusion due to eligibles, treatment violation, noncompliance, or the like holds a promise to have a noise reduction in trial results and make the trial sensitive, but this procedure may also produce a reverse effect due to a decrease in the number of subjects. The argument is strongly urged that the meaning of randomization at the outset is lost when rejections are
large in number. Moreover, when the subjects are homogenized, the nature of population from which conclusions should be drawn becomes more clear. Practically, however, homogenizing the subjects may also decrease the range of application.

17. Exaggeration of pharmacological action

How to handle the signs, symptoms, and other findings anticipated from the undue pharmacological action of an investigational drug should be determined beforehand.

18. Flow chart

It is desirable that a flow chart be used to represent the details of a trial at each stage. In this instance, the state of occurrence of incomplete cases and number of cases included in statistical analysis should be described at each stage.

19. Description of protocol violation

If any important matter of the protocol is violated by a subject or a physician-in-charge, this event should be described, together with the influence of the violation in the widest possible range.

20. Adjustment of background factor bias

In treated groups among which important background factors (prognostic factors) related closely to prognosis are not evenly distributed, analysis results may be distorted when statistical analysis is made without consideration for the effect of prognostic factors. In clinical trials, random allocation is used where possible in order to evenly distribute the subjects among the treated groups with respect to known and unknown prognostic factors. Random allocation is the best means of assuring compatibility among treated group and, even when made without respect to the effect of prognostic factors, analysis produces appropriate comparative results in the great majority of cases. In trials with insufficient numbers of subjects per group, however, prognostic factor bias occurs accidentally among treated groups even after random allocation has been carried out, so that fears may be aroused as to the validity of simple comparison of drug effect differences. When the treated groups differ from one another in the distribution of prognostic factors clearly related closely to prognosis, appropriate statistical adjustment is also necessary as extreme measures.

When the statistical adjustment of prognostic factor bias is made in verificatory clinical trials, the main prognostic factors must be examined for bias prior to the analysis of drug effect differences. In this instance, since the test of prognostic factors themselves is for information, it is important to carefully examine them in terms of the influence of prognostic factor bias on the analysis results of drug effect
differences. In the great majority of cases, background factors, shown by a significant probability of about 15% must be examined carefully for imbalance. When drug effect differences are analyzed statistically and prognostic factors are then selected by hindsight and adjusted statistically, the drug effect of the investigational drug is verified because the impartialness of results cannot be assured.

Of the methods for the statistical adjustment of prognostic factor bias at the stage of analysis, the most fundamental one may be of the stratificational type. In the method by stratification, after stratificational summing is done for each prognostic factor biased, a common difference in drug effects is statistically inferred on the assumption that the difference is common to all strata. Validity of the statistical model as to the presence of a common drug effect difference for each stratum can be confirmed by stratificational summing which also provides a direct understanding from the viewpoint of clinical medicine. As concrete methods for statistical adjustment, the Mantel-Haenszel method is proposed for binary data and expanded Mantel method for data serially classified in 3 or more categories. In order to examine the premise—common difference in drug effects—(examination of interactions), the Breslow-Day method can be used, for example, for binary data. Continuous data can be adjusted statistically by various methods. In the case of a parametric method, however, it is necessary not only to examine continuous data for interactions, but also to ascertain whether or not hypotheses on which such statistical methods are found are valid.

21. Summary of analysis

The values of z and p and number of subjects in each group should be described, for example, when the Wilcoxon two-sample test (Mann-Whitney U-test) has been used. In the case of the two-sample test, description should be given of the value of t, degrees of freedom, value of p, number of subjects in each group, mean, variance, and estimate for composite variance.

22. Verification of equivalence

When evaluated for drug effects, an investigational drug must be significantly superior to placebo and equal to or more potent than a commercially available positive control drug. Evidently, it is not proper to consider the investigational drug to be "equal" to the positive control just because there is "no significant difference." Absence of significant differences does not assure statistical "equivalence."

Clinical equivalence may be statistically verified as follows, for example, in the case of the rate of effectiveness. In the description given below, the symbol indicates a predetermined level of the clinically allowable difference between an investigational drug and a positive control drug level which varies with disease and drug effect (rough guideline level: 10%).

a) The investigational and positive control drugs are judged clinically equal to each other if the
confidence interval for efficacy differences between the two drugs (90% confidence interval) does not cover the range narrower than $-\Delta$.

b) One-tailed hypothetical testing (significance level: 5%) is done for the null hypothesis that "the effect of the investigational drug is inferior to that of the positive control drug by more than $\Delta$." The two drugs are judged clinically equal to each other if the hypothesis is rejected.

This verification method is intended to assure equivalence by proving that the effect of the investigational drug is not inferior to that of the positive control drug by more than $\Delta$. 