ANNEX

to Recommendation No.
of the Eurasian Economic Commission’s Board
dated 20

GUIDELINES
on the Principles of Applying Biostatistics in Clinical Trials

I. General Provisions

1. These Guidelines provide instructions to sponsors in terms of the design, conduct, analysis, and evaluation of clinical trials of an investigational medicinal product in the context of its clinical development. These Guidelines provide instructions on the work of experts responsible for preparing application summaries or assessing evidence of efficiency and safety, mainly based on the results of clinical trials in later phases of medicinal product development.

2. These Guidelines describe the main statistical principles. They do not address the use of specific statistical procedures or methods. Specific practical steps to ensure proper implementation of the principles are the sponsor's responsibility. Integration of data from various clinical trials is discussed, but is not a primary focus of these Guidelines. Some principles and procedures related to data management or clinical trial monitoring activities are covered in other Guidelines of the Union and are not considered here.

3. The principles presented are applicable to various areas of medicinal product trials. In accordance with the Good Clinical Practice Rules of the Eurasian Economic Union approved by Decision No. 79 of the Eurasian Economic Commission's Council dated November 3, 2016, the actual responsibility for all statistical work associated with clinical trials shall lie with an appropriately qualified and experienced statistician. The role and responsibility of the trial statistician in collaboration with other clinical trial professionals, is to ensure that statistical principles are applied appropriately in clinical trials supporting the medicinal product development. Therefore, the trial statistician should have a combination of education and experience sufficient to implement the principles formulated in these Guidelines.

4. All important characteristics of the design and conduct and the fundamental peculiarities of the proposed statistical analysis for each clinical trial contributing to a marketing application should be clearly specified in a protocol to be written before the clinical trial begins. The extent of compliance with the protocol procedures and the planning of the initial analysis will a priori contribute to the degree of confidence in the final results and conclusions of the trial. The protocol and subsequent amendments thereto should be approved by the personnel in charge of such activities, including the trial statistician. The trial statistician should ensure that the protocol and any amendments thereto cover all relevant statistical aspects clearly and accurately, using appropriate technical terminology.

5. The principles set forth in these Guidelines are applicable to clinical trials conducted in the later phases of medicinal product development, many of which are confirmatory trials of medicinal product efficiency. In addition to efficiency, the primary variable in confirmatory clinical trials may be a safety variable (for example, an adverse event, a laboratory variable, or an electrocardiographic parameter), a pharmacodynamic or pharmacokinetic

variable (as in a confirmatory bioequivalence trial). Moreover, some confirmatory results may be derived from data integrated across various trials, hence, some principles of these Guidelines are applicable to this situation. Although the early phases of medicinal product development consist mainly of clinical trials that are exploratory in nature (paragraphs 16 and 17 of these Guidelines), statistical principles set forth herein are also applicable to them. Therefore, the principles set forth in these Guidelines should, as far as possible, be applied to all phases of clinical development.

6. The approaches described in these Guidelines enable minimizing bias and maximizing the accuracy (precision) of statistical estimates. Nevertheless, the potential sources of bias should be identified as completely as possible in order to try to eliminate them.

The presence of bias may seriously compromise the ability to obtain valid conclusions from the results of clinical trials.

7. Some sources of bias are due to the trial design, for example, the allocation of subjects to trial groups where patients with low risk systematically get into one of the groups. Other sources of bias arise during the conduct and analysis of a clinical trial. For example, protocol violations and exclusion of subjects from analysis based on the knowledge of a specific outcome for a particular subject are possible sources of bias that may affect the accurate assessment of the treatment effect. As bias may occur for implicit or unknown reasons, and its impact cannot be measured directly, it is necessary to evaluate the robustness of the results obtained and main conclusions of the trial. Robustness is understood as the degree of sensitivity of the final conclusions to various data limitations, statistical assumptions and approaches to data analysis. The robustness of the results obtained and final conclusions implies that the analysis based on alternative assumptions or analytical approaches does not significantly influence on the treatment effect

and the primary conclusions of the trial. The interpretation of statistical measures of uncertainty of the treatment effect and treatment comparisons should involve consideration of the potential contribution of bias to the p-value, confidence interval and conclusions.

8. As the predominant approaches to the planning and analysis of clinical trials have historically been based on frequentist statistical methods, these Guidelines mainly refer to the use of frequentist methods of statistical analysis when discussing verification of hypotheses and (or) confidence intervals. At the same time, other approaches to statistical analysis and hypothesis verification that are not based on frequentist methods may be used, such as the Bayesian and other approaches, if the reasons for their use are clearly defined, and the results and conclusions obtained are sufficiently robust.

II. Definitions

9. For the purposes of these Guidelines, concepts shall be used with their following meanings:

«bayesian approaches» – mean approaches to data analysis allowing to obtain a posteriori probability distribution of a certain parameter (for example, treatment effect) on the basis of the observed data and a priori probability distribution of such parameter, and using the obtained a posteriori distribution as a basis for statistical inference;

«safety and tolerability» – the safety of a medicinal product is characterized by the medical risk to the subject, usually assessed in a clinical trial by laboratory tests (including clinical biochemistry and hematology), vital signs, clinical adverse events (diseases, symptoms and complaints), and other special safety tests (for example, electrocardiograms, ophthalmologic examination). The tolerability of a medicinal product is characterized by the

degree to which explicit adverse reactions can be tolerated by the trial subject;

«interaction (qualitative and quantitative)» – the situation in which a treatment contrast (for example, difference between investigational medicinal product and control treatment) is dependent on another factor (for example, a clinical center that is involved in medicinal product trials). A quantitative interaction refers to the case where the magnitude of the contrast differs for different levels of the factor, whereas for a qualitative interaction the direction of the contrast differs for at least one level of the factor;

«dropout» – means a subject of a clinical trial who for any reason failed to continue participating in the trial until the last visit required of him/her by the trial protocol;

«generalizability, generalization» – means the extent to which the data of a clinical trial may be reliably extrapolated from the subjects who participated in the trial to a broader patient population and a broader group of clinical conditions for the medicinal product use;

«double dummy» – means a technique for retaining the blind when distributing medicinal product supplies for treatment in a clinical trial, when the two treatments cannot be made identical. Medicinal product supplies are prepared for Treatment A (active medicinal product and indistinguishable placebo) and for Treatment B (active medicinal product and indistinguishable placebo). Then subjects of the clinical trial take two sets of supplies: either active medicinal product A and placebo B, or placebo A and active medicinal product B;

«non-inferiority trial» – means a trial with the main purpose to show that the response to the investigational treatment is not clinically significantly inferior to a comparative treatment (active or placebo);

«superiority trial» – means a trial with the main purpose to show that the response to the investigational treatment is superior to that to a comparative treatment (active or placebo control);

«equivalence trial» – means a trial with the main purpose to show that the response to two or more treatments differs by a value that is clinically insignificant. This is usually demonstrated by showing that the true treatment difference is between a lower and an upper equivalence margin of clinically acceptable differences;

«meta-analysis» – is the formal evaluation of the quantitative evidence from two or more trials aiming at the same question. This most commonly involves the statistical combination of summary statistics from various trials, but the concept is sometimes also used to refer to the combination of raw data;

«inter-rater reliability» – means the property of obtaining equivalent results when different experts on different occasions are involved;

«blind review» – means the review and assessment of data during the period of time between the trial completion (the last observation on the last subject) and the blind breaking, with the purpose of finalizing the analysis plan;

«global assessment variable» – means a single variable, usually a scale of ordered categorical assessments (ratings), which integrates objective indicators and the investigator's subjective overall impression about the state or change in state of the trial subject;

«statistical analysis plan» – means a separate document that contains a detailed and described in a more specialized language (technical terms) statement of the main features of the analysis described in the protocol, and includes detailed procedures for carrying out the statistical analysis of the primary and secondary variables and other data;

«confirmatory clinical trial» – means a clinical trial where a previously formulated hypothesis is verified and which is adequately controlled;

«full analysis set» – means the set of trial subjects that is as close as possible to the theoretical concept (the «intention-to-treat» principle), and derived from the set of all randomized trial subjects by way of minimal and justified elimination of subjects from the trial;

«preferred and included terms» – mean the lowest level of vocabulary terms in the hierarchical dictionary, using which the investigator's description is encoded, while the preferred term refers to the level of grouping of included terms typically used in reporting within the framework of a clinical trial for the purpose of further calculating the phenomenon occurrence frequency (for example, the investigator's text «Pain in the left arm» may be encoded by the included term «Joint pain», which is reported at the preferred term level as «Arthralgia»);

«intention-to-treat principle» – means the principle establishing that the effect of a treatment tactic can be best assessed by analysis depending on the initial intentions to treat the subject (i.e. the planned treatment regimen) rather than the actual treatment given. Consequently, the trial subjects allocated to a treatment group should be followed up, assessed, and analyzed as part of the initial group for the corresponding treatment irrespective of their actual commitment to the initially planned course of treatment;

«per protocol set» principle» – (valid cases, efficiency sample, evaluable subjects sample) means the principle establishing that the effect of treatment tactics is best assessed by analyzing the actual treatment regimen, that is, by analyzing the set of data generated by the subset of subjects who complied with the clinical protocol requirements sufficiently to ensure that these data would be likely to exhibit the effects of treatment according to the scientific model of the trial. Compliance with the requirements of the clinical

trial protocol includes such conditions as the subject of the trial reaching the necessary degree of treatment, the availability of measurement results for the subject and the absence of serious protocol violations by this trial subject;

«interim analysis» – means any analysis intended to compare treatment groups with respect to efficiency or safety at any time prior to the formal completion of a clinical trial;

«bias (statistical and operational)» – means the systematic tendency of any factors associated with the design, conduct, analysis and evaluation of the results of a clinical trial to make the estimate of a treatment effect deviate from its true value. Bias introduced through deviations in trial conduct is referred to as operational bias, the other bias are referred to as statistical bias;

«content validity» – means the extent to which the estimates of a variable (for example, a rating scale) corresponds to the real measurable range of the type of treatment in a clinical trial;

«trial statistician» – means a statistician who has a combination of education, training and experience sufficient to implement the principles set forth in these Guidelines and who is responsible for the statistical aspects of the trial;

«surrogate variable» – means a variable that allows measuring the effect of the treatment in an indirect way where a direct measurement of the clinical effect is impossible or not feasible;

«characteristic absence» – means the value of a variable that unambiguously determines either the absence of a certain state (for example, confirms the absence of a disease) or the confirmed lack of knowledge about the indicator (for example, the planned answer «I don’t know» or «no information» along with the answers «yes» and «no»). In contrast to the missing value, which may be made up in different ways, the characteristic absence uniquely determines the value of the indicator;

«frequentist (statistical) methods» – mean statistical methods, such as significance tests and confidence intervals, which may be interpreted in terms of the frequency of certain outcomes occurring in hypothetical repeated realizations of the same experimental situation;

«treatment effect» – means an effect attributed to a treatment in a clinical trial. In most clinical trials, the treatment effect of interest is a comparison (or contrast) of two or more treatments;

«treatment emergent» – means an event that emerges during treatment (having been absent prior to treatment) or worsens relative to the pretreatment state.

Other concepts used in these Guidelines shall have the meanings set out in the Good Clinical Practice Rules of the Eurasian Economic Union approved by Decision No. 79 of the Eurasian Economic Commission's Council dated November 3, 2016 (hereinafter - the Good Clinical Practice Rules).

III. Statistical principles in drawing up the program for clinical development of a medicinal product

1. Trial Context

Development Plan

10. The process of clinical development of a new medicinal product is aimed at ascertaining whether there is a dose range and dosage regimen at which the simultaneous safety and efficiency of the medicinal product may be shown to the extent that the risk-benefit ratio is acceptable as well as identifying specific subjects of the trial who may benefit from the medicinal product, and the specific indications for its use.

11. Achieving these global objectives usually requires an ordered clinical development program in the form of a set of clinical trials, each

having its own specific objectives, as specified in the Guidelines on General Considerations for Clinical Trials (Annex to Recommendation No. 11 of the Eurasian Economic Commission's Board dated July 17, 2018). The clinical development program should be stated in one or several clinical trial protocols with appropriate time points for decision-making and sufficient flexibility to evaluate them, enabling to modify the clinical development program as the information accumulates.

12. A marketing authorization application for a medicinal product should clearly describe the main content of such plans, and the contribution made by each trial. The interpretation and assessment of evidence obtained from the total program of trials involves the synthesis of evidence from individual trials (as described in Subsection 2 of Section VIII of these Guidelines). To facilitate the synthesis of evidence, it is necessary to use standard characteristics of trials (for example, dictionaries of medical terms, wording and terms of assessment of basic measurements, procedures for processing deviations from the protocol, etc.). If medical questions are covered in several trials, information synthesis techniques such as a statistical summary, overview, or meta-analysis may be informative. Where possible, these questions should be presented in the analysis plan so that the relevant trials are identified and any necessary common features of their designs are specified in advance. Other significant statistical issues (if any) that may affect a number of trials in a common plan should be addressed in this plan.

Confirmatory Clinical Trial

13. Confirmatory clinical trials are required to provide firm evidence of the efficiency or safety of an investigational medicinal product. In such trials, the key hypothesis of interest follows directly from the trial's main objective; it is predefined and constitutes the hypothesis that is subsequently verified

when the trial is completed. In a confirmatory trial, it is equally important to estimate with due accuracy (precision) the size of the effects associated with the treatment of interest and to compare these effects with their clinical significance.

14. Confirmatory clinical trials are intended to provide clear evidence to support the relevant hypotheses (statements); therefore, the extent of adherence to protocols and standard operating procedures is particularly important. All unavoidable changes should be explained and documented with a simultaneous assessment of their effect. The rationale of the design of each such trial and of other important statistical aspects, such as the principal features of the planned analysis, should be set out in the clinical trial protocol. Each trial should address a strictly limited range of issues.

15. Firm evidence to support the relevant hypotheses implies that the results of the confirmatory trials will demonstrate that the investigational medicinal product has clinical benefits. In this regard, the confirmatory clinical trials should be sufficient to answer clearly and unambiguously each clinical question that is most consistent with the hypotheses about the efficiency or safety of the investigational medicinal product. In addition, it is necessary to provide an explanation and a basis for generalizing the results to the target patient population; it may also affect the number and type of clinical centers (such as specialized or general practice) and (or) trials required. The results of the confirmatory trial(s) should be robust. In some cases, evidence based on the results of a single confirmatory trial may be significant enough.

Exploratory Clinical Trial

16. The scientific rationale and design of confirmatory clinical trials should be based on previous clinical work carried out as part of a series of

exploratory trials. Like all other clinical trials, these exploratory trials should have clear and specific objectives. At the same time, in contrast to confirmatory trials, their objectives may not always lead to a simple verification of predefined hypotheses. In addition, exploratory clinical trials often require a more flexible approach to design so that changes can be made to the trial plan as the experience accumulates. Their analysis may consist in exploratory assessment of data. Hypothesis verifications may be carried out, but the choice of hypothesis may directly depend on the data obtained. Such trials cannot constitute the ground for the formal proof of efficiency, although they may contribute to the total body of relevant evidence.

17. Any individual trial may include both confirmatory and exploratory aspects. For example, in most confirmatory trials the data are also subjected to exploratory analyses which serve as a basis for explaining or supporting their main results as well as elaborating new hypotheses for subsequent studies. The clinical trial protocol should make a clear distinction between the aspects of a trial to be used for supporting evidence and the aspects to constitute the ground for the exploratory analysis.

2. Scope of Trials

Population of Clinical Trial

18. In the earlier phases of medicinal product development, the choice of subjects for a clinical trial may be greatly influenced by the wish to maximize the chance of observing specific clinical effects of interest, whereby subjects may come from a very narrow subgroup of the entire patient population, for which the medicinal product may eventually be prescribed. Along with that, by the time the confirmatory trials begin, the trial subjects should more closely correspond to the planned target population for treatment. In such trials, it is usually appropriate to minimize the rigidity of

inclusion and non-inclusion criteria within the target population, while maintaining sufficient population uniformity necessary for accurate (precision) assessment of treatment effects. No individual clinical trial can be expected to be totally representative of population of the treatment future users because of the possible impact of geographical location, the time of the clinical trial, the medical practices of the particular investigators and clinics, etc. At the same time, the influence of such factors should be minimized wherever possible, and then analyzed when interpreting the trial results.

Primary and Secondary Variables of Clinical Trial

19. The primary variable («target» variable, primary endpoint) should be the variable capable of providing the most clinically relevant and convincing evidence directly related to the primary objective of the trial. As a rule, there should be only one primary variable. Usually, the efficiency variable is chosen as the primary objective of most confirmatory clinical trials is to obtain scientific evidence regarding efficiency. Safety and (or) tolerability, which is always an important aspect of the trial, may sometimes be the primary variable. Parameters related to quality of life and health economics may also be primary variables.

20. The selection of the primary variable should reflect the accepted norms and standards in the relevant field of medicine. It is recommended to use a reliable and validated variable, data on which was obtained in earlier trials or published in the scientific literature. Sufficient evidence should be provided that the primary variable may be a valid and reliable measure of clinically significant and important treatment benefit in the patient population described by the criteria for the inclusion and non-inclusion of subjects in the trial. The primary variable should be used to estimate the sample size

(Subsection 5 of Section IV of these Guidelines), unless a different approach is justified.

21. In many cases, the approach to assessing outcomes may not be straightforward and should be carefully regulated (for example, it is unacceptable to choose mortality as a primary variable without further clarification; mortality may be assessed by comparing proportions of alive at fixed time points or by comparing overall distributions of survival times during the specified interval). Another common example is a recurring event; the measure of treatment effect may be a simple dichotomous variable (any occurrence during a specified period of time), time to first occurrence, rate of occurrence (number of events per observation time unit), etc.

22. The assessment of functional status of the trial subject over time when studying treatment in chronic diseases causes difficulties in selecting the primary variable. There are many possible approaches, such as comparisons of the assessments made at the beginning and end of the interval of observation; comparison of regression straight slopes calculated from all estimates over the entire observation interval; comparisons of the proportions of subjects with indicators exceeding or declining beyond a specified threshold, or comparisons based on methods for repeated measurements of data. To avoid the concerns of multiplicity when testing hypotheses that arise in connection with post hoc comparisons, it is critical to provide in the clinical trial protocol the precise definition of the primary variable as it will be used in the statistical analysis. The clinical significance of the specific selected primary variable and the validity of the associated measurement procedures should also be described and justified in the clinical trial protocol.

23. The primary variable should be specified in the clinical trial protocol, along with the rationale for its selection. Redefinition of the primary variable after blind breaking will almost always be unacceptable, since the

resulting biases are difficult to assess. If the clinical effect defined by the main objective will be measured by several methods, the protocol of the clinical trial should identify one of the measurement methods as the primary variable on the basis of clinical significance, importance, objectivity, and (or) other relevant characteristics, whenever such selection is feasible.

24. The secondary variables are either supportive measurements related to the primary objective of the clinical trial, or measurements of effects associated with the secondary objectives. They should also be predefined in the protocol of the clinical trial, and their relative importance and contribution to the interpretation of the trial results should be explained. The number of secondary variables should be limited and consistent with the limited number of questions to be answered in the trial.

Composite Variables of Clinical Trial

25. If a single primary variable cannot be selected from the set of parameters associated with the main objective, another possible strategy is to integrate or combine the multiple parameters into a single «composite» variable, using a predefined algorithm. In some cases, the primary variable is indeed a combination of several clinical parameters (for example, the rating scales used in arthritis, psychiatric disorders, etc.). This approach addresses the multiplicity problem without requiring adjustment to the Type I error.

26. The method of combining several measurements should be given in the clinical trial protocol, and the interpretation of the resulting scale should be provided in terms of the size of the clinically significant benefit of the treatment. If a composite variable is used as a primary variable, the components of this variable may sometimes be analyzed separately, if it is clinically justified and validated. If a rating scale is used as a primary variable, it is especially important to consider factors such as content validity,

inter- and intra-rater reliability, flexibility (response time) and sensitivity (size of change) in relation to changes in the severity of the disease.

Global Assessment Variables of Clinical Trial

27. In some cases (usually in neurology and psychiatry), global assessment variables are developed to measure the overall safety, overall efficiency, and (or) overall usefulness of a treatment. A variable of this type integrates objective variables and the general impression of the investigator about the state or change in the state of the subject of trial and usually represents a scale of ordered categorical ratings.

28. Global assessment variables generally have a subjective component. When a global assessment variable is used as a primary or secondary variable, further details of the scale should be included in the protocol with respect to:

a) the significance of the scale for assessing the primary objective of the trial;

b) the basis for recognizing the validity and reliability of the scale;

c) the methods of handling data received from an individual subject to assign him/her to a specific category of the scale;

d) the methods for assigning subjects with missing data to a certain category (response option) of a measurement scale or other way of evaluating them.

29. If the investigator considers objective variables during the global assessment, then such variables should be considered as additional primary or, at least, important secondary variables.

30. The global assessment of usefulness combines the components of both benefit and risk that arise from treatment and reflects the decision-making process by the treating physician, who should weigh benefit and risk

when deciding on the product prescription. The disadvantage of global usefulness assessment variables is that their use in some cases can lead to the recognition of two medicinal products as equivalent, despite having very different profiles of beneficial and adverse reactions. For example, recognizing the global usefulness of an investigated treatment as equivalent or superior to the usefulness of the alternative treatment can mask the fact that the investigated treatment has little (or no) efficiency with fewer adverse reactions. In this regard, the global usefulness variable is not recommended as a primary variable. If global usefulness is specified as primary, specific outcomes of efficiency and safety should be considered separately as additional primary variables.

Multiple Primary Variables of Clinical Trial

31. In certain situations, it is desirable to use several primary variables, each of which (or a subset of which) could be sufficient to cover the range of effects of therapies. The planned manner of interpretation of this type of evidence should be thoroughly described. It should be clear whether the trial objective is achieved with the proven effect on at least one of the variables, on a certain number of them, or on all variables. The primary hypothesis or hypotheses and estimates of interest (for example, mean, percentage, distribution) should be clearly stated in relation to the selected primary variables and the described approach to the statistical conclusion (defining the statistical conclusion). Due to the potential problem of multiplicity, it is necessary to explain the effect of the conducted comparisons on the Type I error (Subsection 6 of Section VI of these Guidelines), the method for controlling Type I error should be given in the clinical trial protocol. The extent of inter-correlation between the proposed primary variables may be analyzed when assessing the effect of multiple comparisons on a Type I error.

If the purpose of the trial is to demonstrate the effect on all of the selected primary variables, then there is no need to introduce adjustment for the Type I error, however, the effect of the multiplicity of comparisons on Type II error and sample size should be carefully analyzed.

Surrogate Variables of Clinical Trial

32. If there is no practical possibility of direct assessment of the clinical benefit to the subject of the trial by monitoring the actual clinical efficiency, indirect criteria (surrogate variables) may be considered. Commonly accepted surrogate variables are used in a number of indications where they are believed to be reliable predictors of clinical benefit. There are two main concerns when using any proposed surrogate variable. First, it may not be a true prognostic factor of the clinical outcome of interest. For example, it may measure treatment activity associated with one specific pharmacological mechanism, but may not provide full information on the range of actions and key effects of the treatment, whether positive or negative. There are many examples where treatments showing a pronounced beneficial effect on a proposed variable were ultimately be detrimental to the subjects' clinical outcome, and conversely, there are examples of treatments conferring clinical benefit without measurable impact on proposed surrogate variables. Second, proposed surrogate variables may not correlate with clinical benefit value that could be further compared directly against adverse reactions. Statistical criteria for the validation of surrogate variables are proposed but the experience with their use is relatively limited. In practice, the reliability of a surrogate point depends on the following factors:

the biological plausibility of the relationship between clinical efficiency and surrogate point;

the demonstration in epidemiological trials of the prognostic value of the variable selected as a surrogate point for assessing the clinical outcome;

the evidence based on previous clinical trials that treatment effects on the surrogate point correlates with the treatment effect on the clinical outcome.

Relationships between clinical efficiency and surrogate points for one medicinal product do not necessarily apply to a medicinal product with a different mechanism of action when treating the same disease.

Categorized Variables of Clinical Trial

33. Where relevant, an approach with a dichotomous categorization of the studied variable may be used, or other categorization of continuous or ordinal variables may be implemented in a clinical trial. The most common dichotomous categorization is the use of «success /failure» or «response/no response» criteria. Dichotomous criteria require:

precise definition, for example, specifying the minimum percentage of improvement (compared to the baseline) if used to evaluate a continuous variable;

specifying a value corresponding to a certain threshold level, exceeding which allows you to set the appropriate category (for example, “good”), in the case of an ordinal rating scale. An example of a threshold level is the use of a 90 mmHg diastolic pressure regarding which dichotomous categorization of the level of pressure reduction (“response to treatment/no response to treatment”) takes place.

Variable categorization should be used if it may help to evaluate the clinical significance of the treatment. The criteria for categorization should be predefined and specified in the protocol of a clinical trial, as knowledge of trial results could be a source of bias when choosing categorization criteria.

As categorization of variables normally implies a loss of information during the categorization process, the consequence will be a reduction of statistical power in the data analysis; this should be accounted for in the sample size calculation.

3. Planning Methods to Minimize Bias

34. The most important planning methods to minimize bias in clinical trials are blinding and randomization, which should be standard features of most controlled clinical trials to be included in a marketing application of a medicinal product. Most such trials use a double-blind method in which treatment supplies are prepacked and delivered to the trial center(s) labeled only with the subject number and the treatment period, so that no one involved in the trial is aware of the treatment assigned to a specific subject of the trial (not even as a code letter). This approach is described in paragraphs 38-43 of these Guidelines.

35. Bias can also be reduced at the design stage of a clinical trial by including procedures in the clinical trial protocol aimed at minimizing all expected deviations in trial conduct that might impair the satisfactory analysis, including various types of protocol violations, withdrawals and missing values. The clinical trial protocol should provide ways both to reduce the frequency of such problems and to handle the consequences of such problems that occur in the data analysis stage.

Blinding (Masking) Method

36. Blinding (masking) method is intended to limit the occurrence of conscious and unconscious bias in the conduct and interpretation of a clinical trial related to the influence of the initial knowledge of studied treatment in each group of subjects on:

the recruitment of subjects in clinical trial; the allocation of trial subjects to groups;

the subsequent medical care to subjects;

the attitudes of subjects to the studied treatments;

the assessment of end-points;

the handling data from withdrawals;

the exclusion of data from analysis;

other aspects of the clinical trial and its statistical evaluation.

37. The fundamental aim of the blinding (masking) method is to prevent identification of the treatments until all opportunities for bias occurrence have passed.

38. A double-blind trial is a trial in which neither the subject nor any of the investigator or sponsor staff involved in the treatment or clinical evaluation of the subjects (including anyone who determines subject eligibility, evaluates endpoints, or verifies compliance with the protocol of clinical trial) are aware of the treatment received. This level of blinding of the obtained raw data is maintained throughout the trial, and only when the data are cleaned to an acceptable level of quality, the appropriate personnel will be unblinded. If any of the sponsor staff who are not involved in the treatment or clinical evaluation of the subjects are required to be unblinded to the treatment code (for example, bioanalysts, auditors, persons involved in serious adverse event reporting), the sponsor should have appropriate standard operating procedures to protect against improper distribution of treatment codes. In a single blind trial the investigator and (or) his staff are aware of the treatment but the subject is not, or vice versa. In an open-label trial, the identity of treatment is known to all. The double-blind trial is the optimal approach. It requires that the treatments used in the trial cannot be distinguished (by appearance, taste, etc.) either before or during

administration, and that the blinding is maintained properly during the whole trial.

39. Difficulties in reaching the maximum double-blind level may arise due to the following factors: treatment nature may be completely different, for example, surgery and medicamentous therapy; two medicinal products may have different formulations and, although they could be made indistinguishable by the use of capsules, changing the formulation might also change the pharmacokinetic and (or) pharmacodynamic properties and hence require that bioequivalence of the formulations be demonstrated; the daily pattern of administration of two treatments may differ. One way of achieving double-blinding under these circumstances is to use a double-dummy method. This method may sometimes force a medicinal product administration pattern that is unusual that it may adversely affect the motivation and compliance of subjects to the clinical trial protocol. Ethical problems (for example, the use of imitation of surgical procedure to provide the necessary degree of blinding (masking) of treatment types). None the above cases shall constitute the ground for refusing to use the double-blinding method on a formal basis.

40. The double-blind nature of some clinical trials may be partially compromised by the clear effects induced by the treatment. In such cases, the reliability of blinding may be improved by blinding investigators and the sponsor staff involved in the trial with respect to the results of certain tests (for example, certain laboratory parameters). In trials where the unique or specific treatment effects may lead to the unblinding individual patients, it is necessary to provide approaches to minimize bias similar to those used in open-label trials.

41. If a double-blind trial is not feasible, then the single blind option should be considered. In some cases, only an open-label trial is practically or

ethically possible. Single blind and open-label trials provide additional flexibility, but it is particularly important that the investigator's knowledge of the next treatment should not influence the decision to enter the subject; this decision should precede knowledge of the randomized treatment. In the case of such trials, it is advisable to provide a centralized randomization method, for example, telephone randomization, to organize the assignment of randomized treatment. In addition, clinical assessment should be carried out by medical staff who are not involved in treating subjects and who remain blind to treatment.

42. In single blind and open-label trials, every effort should be made to minimize the various known sources of bias, and the primary variables should be as objective as possible. The clinical trial protocol should include the reasons for choosing the type of blinding along with measures taken to minimize bias using other methods. For example, the sponsor should have adequate standard operating procedures to ensure that access to the treatment code is appropriately restricted during the process of database cleaning prior to being submitted for analysis.

43. Blinding (for a single subject) should be broken only when knowledge of the treatment assignment is deemed obligatory by the subject's physician for the subject's care. Any intentional or unintentional blind breaking should be reported and explained at the end of the trial, regardless of the reasons for its occurrence. The procedure and timing for revealing the treatment assignment codes should be documented.

Randomization

44. Randomization introduces a deliberate element of chance into the assignment of treatments to subjects in a clinical trial. During subsequent analysis of the trial data, it provides a firm statistical basis for the quantitative

evaluation of the evidence related to treatment effects. It also helps obtain treatment groups with similar distributions of prognostic factors, both known and unknown. Along with blinding, the randomization helps avoid possible bias when selecting subjects and allocating them to groups, which occur due to the predictability of treatments prescribed.

45. The randomization schedule of a clinical trial documents the random allocation of treatments to subjects. In the simplest case it is a sequential list of treatments (or treatment sequences in a crossover trial) or corresponding codes by subject number. The logistics of some trials, such as those having a screening phase, may make randomization more complicated, but the specific and preplanned assignment of treatment or sequence of treatments to the subject should be described clearly. Different trial designs will require different procedures for generating randomization schedules. The randomization schedule should be reproducible where required.

46. Despite the fact that single or unrestricted randomization is an acceptable approach, some advantages can usually be achieved by block randomization (randomizing subjects in blocks). It helps to increase the comparability of the treatment groups, particularly if the subject characteristics may change over time, for example, as a result of changes in recruitment policy. It also provides a better guarantee that the treatment groups will be of nearly equal size. In crossover trials, it allows to achieve balanced designs characterized by greater efficiency and easier interpretation of the results. Block lengths should be selected with care, these blocks would be short enough to limit possible imbalance, but long enough to avoid predictability at the end of the sequence in the block.

47. Investigators and other relevant staff involved in the trial should, as a rule, be blind with respect to block length; the same goal can be achieved using two or more block lengths randomly selected for each block

(theoretically, in a double-blind trial predictability does not matter, but the pharmacological effects of medicinal products may provide the opportunity for making assumptions based on studying common factors).

48. In multicenter trials, the randomization procedures should be organized centrally. It is recommended to have a separate random scheme for each center, i.e. with a view to stratify by center or to allocate several whole blocks to each center. More generally, stratification by important prognostic factors, measured at baseline, that is, before the start of the treatment (for example, the severity of the disease, age, gender, etc.) may be a valuable way to improve the balance of the allocation within the strata; it has greater potential benefit in small trials.

49. The use of more than two or three stratification factors is rarely required; it is difficult to achieve a balance in this situation, besides, it is logistically troublesome. The use of a dynamic allocation procedure (see below) may help to achieve balance across a number of stratification factors simultaneously, provided the rest of the trial procedures can be adapted to a similar approach. Factors for which randomization has been stratified should be considered in a subsequent analysis.

50. The next subject to be randomized into a trial should always receive the treatment corresponding to the next free number in the appropriate randomization schedule (in the respective stratum, if randomization is stratified). The appropriate number and associated treatment for the next subject should only be allocated when entry of that subject to the randomized part of the trial has been confirmed.

51. The trial protocol should not include such a detailed description of randomization that would allow predicting the allocated group (for example, specifying the block length). The sponsor or an independent party is required to keep the randomization schedule itself private, ensuring that blindness is

properly maintained throughout the trial. Access to the randomization schedule during the trial should take into account the possibility that, in an emergency, the blind may have to be broken for any subject. The procedure to be used, the necessary documentation, and the subsequent treatment and subject assessment should be described in the clinical trial protocol.

52. Dynamic allocation is an alternative procedure to randomization, in which the allocation of treatment to a subject is influenced by the current balance of allocated treatments and, in a stratified trial, by the stratum to which the subject belongs and the balance within that stratum. It is necessary to avoid dynamic allocation procedures based on predefined (deterministic) algorithms, and to introduce an appropriate element of randomization for each treatment allocation. Maximum efforts should be made to preserve the double-blinding of the trial. For example, knowledge of the treatment code may be restricted to a central trial office from where the dynamic allocation is controlled, generally through telephone contact. This in turn permits to perform additional checks on compliance with eligibility criteria and arranges entry into the trial, features that can be valuable for certain types of multicenter trials. Then it may be possible to follow the usual system of preliminary packaging and labeling of medicinal product supplies for double-blind trials, however, the order of their use is no longer sequential. It is desirable to use appropriate computer algorithms so that the personnel at the central trial office remains blinded to the treatment code. The complexity of the logistics and potential impact on the analysis should be carefully evaluated when considering dynamic allocation.

IV. Trial Design Issues

1. Design Choice

Parallel Group Design

53. The most common design for confirmatory clinical trials is the parallel group design in which subjects are randomized to one of two or more arms, and each arm is allocated a different treatment. These treatments include the investigational product at one or more doses, and one or more control treatments, such as placebo and (or) an active comparator.

54. The assumptions underlying this design are less complex than for most other designs. At the same time, like other designs, there may be additional features of the trial that complicate the analysis and interpretation (for example, covariates, repeated measurements over time, interactions between design factors, protocol violations, availability of dropouts and withdrawals).

Crossover Design

55. In the crossover design, each subject is randomized to a sequence of two or more treatments and as a result acts as his own control for treatment comparisons. This simple approach is attractive primarily because it reduces the number of subjects and usually the number of assessments needed to achieve a specific power, sometimes to a large extent. In the simplest 2x2 crossover design, each subject receives two treatments in randomized order in two successive treatment periods, often separated by a washout period. The most common extension of this involves comparing n ($n > 2$) treatments in n periods, with each subject receiving all n treatments. There are numerous variations, such as designs in which each subject receives a subset of n ($n > 2$) treatments, or designs in which treatments are repeated within a subject.

56. Crossover designs have a number of problems that can invalidate their results. The primary difficulty concerns carryover, that is, the residual influence of treatments in subsequent treatment periods. In an additive model, the effect of carryover different between groups will introduce bias into direct

treatment comparisons. In the 2x2 crossover design, the carryover effect cannot be statistically distinguished from the interaction effect between treatment and period, and the significance verification for either of these effects lacks sufficient power because the corresponding contrast is between subjects. This problem is less pronounced when using higher order designs, but it cannot be completely ignored. When the crossover design is used, it is therefore important to avoid carryover. The easiest way to achieve this is to make a careful design choice based on sufficient knowledge of both the disease and the new medicinal product. The studied disease should be chronic and stable. The effects of interest of the medicinal product should develop fully within the treatment period. The washout periods should be sufficiently long for complete reversibility of effect of the medicinal product. The fact that these conditions are likely to be met should be established before the trial based on previous information and data.

57. There are additional problems that require careful attention in crossover trials. The most significant of these are the complications of analysis and interpretation arising from the loss of subjects. Besides, the potential carryover makes it difficult to assign adverse events that occur in subsequent treatment periods to the appropriate treatment. These and other issues are described in the Guidelines for the Selection of Doses of Medicinal Products (Annex to Recommendation No. 8 of the Eurasian Economic Commission’s Board dated March 12, 2019). The crossover design should generally be restricted to situations where small loss of subjects is expected.

58. A common, and generally satisfactory, use of the 2x2 crossover design is to demonstrate the bioequivalence of two production formulations of the same medicinal product. In this particular application in healthy volunteers, carryover effect on the relevant pharmacokinetic variable is less likely to occur if the wash-out time between the two periods is sufficiently

long. However, it is still important to verify this assumption during analysis based on the data obtained, for example, by demonstrating that no medicinal product is detected at the start of each period.

Factorial Designs

59. In a factorial design, two or more treatments are evaluated simultaneously through the use of varying combinations of the treatments. The simplest example is the 2x2 factorial design in which subjects are randomly allocated to one of the four possible combinations of two treatments, A and B:

treatment A alone;

treatment B alone;

treatments A and B both;

neither A, nor B.

60. In many cases, this design is used for the specific purpose of examining the interaction of treatments A and B. The statistical test of interaction may not be powerful enough to detect interaction if the sample size was calculated based on the test for main effects. This is important to consider when a similar design is used to study the joint effects of A and B, in particular, if the treatments may be used together.

61. Another important use of the factorial design is to establish the dose-response characteristics with the simultaneous use of treatments C and D, especially if the efficiency of each monotherapy was established at a certain dose in previous trials. A number (m) of doses of C is selected, usually including a zero dose (placebo), and a similar number (n) of doses of D. The full design then consists of m x n treatment groups, each of which receives a different combination of doses of C and D. The resulting estimate of the response surface may then be used to identify the appropriate

combination of doses of C and D for clinical use, as described in the Guidelines for the Selection of Doses of Medicinal Products.

62. In some cases, a 2x2-crossover design may be used to make efficient use of clinical trial subjects by evaluating the efficiency of the two treatments with the same number of subjects that would be required to evaluate the efficiency of any of them. This strategy has proved to be of particular value for very large mortality trials. The efficiency and validity of this approach depends upon the absence of interaction between treatments A and B so that the effects of A and B on the primary efficiency variables correspond to the additive model, and therefore effect A is actually identical, regardless of whether it is complementary to effect B, or studied separately. Similar to the crossover design of the trial, evidence that this condition will be met should be obtained prior to the trial based on previous information and data.

2. Multicenter Trials

63. Multicenter trials are carried out for two main reasons. First, a multicenter trial is an acceptable way of evaluating a new medicinal product more efficiently; under some circumstances, it may be the only practical way to recruit subjects sufficient to satisfy the trial objective within a reasonable time. Multicenter trials of this nature may, in principle, be carried out at any stage of clinical development. There may be several centers with a large number of subjects per each or, in the case of a rare disease, there may be a large number of centers with very few subjects per center. Second, a trial may be designed as a multicenter (and multi-investigator) trial mainly for a more reliable basis for the subsequent generalization of the results. This is due to the possibility of recruiting subjects from a wider population and of administering a medicinal product in a wider range of clinical conditions.

Thus, a situation is experimentally modeled that is more typical to future conditions of use. In this case, the participation of a large number of investigators also contributes to a wider range of clinical judgments regarding the value of the medicinal product. This type of trial would be a confirmatory trial in the later phases of medicinal product development and would include a large number of investigators and centers. The trial might sometimes be conducted in different countries to facilitate generalizability even further.

64. For the correct interpretation and extrapolation of a multicenter trial, the method of the trial conduct according to the protocol should be clear and similar in all centers. Moreover, the standard sample size and power calculations depend on the assumption that the differences between the compared treatments in the centers are unbiased estimates of the same quantity. It is necessary to design a common protocol and to conduct the trial bearing in mind this assumption. Procedures should be standardized as completely as possible. Variation of evaluation criteria and schemes can be reduced by investigator meetings, by the training of personnel prior to the trial, and by careful monitoring during the trial. High-quality design should, as a rule, be aimed to achieve the same allocations of subjects by treatments within each center and good trial management should help to achieve this design goal. Trials that limit excessive variation in the number of subjects per center, and trials that exclude the option of a few very small centers, have advantages if subsequently it becomes necessary to take into account the heterogeneity of the effect of treatment between the centers, since there is less difference between different weighted estimates of the treatment effect (this does not apply to trials where all centers are very small or in which the center does not appear in the analysis). Failure to take these precautions, combined with doubts about the homogeneity of the results, may, in severe cases, reduce the value of a multicenter trial to such a degree that it cannot be

considered as a source of convincing evidence of hypotheses put forward by sponsors.

65. In the simplest multicenter trial, each investigator is responsible for the subjects recruited at one hospital, so that center is identified uniquely by either investigator or hospital. However, the situation is usually more complicated in clinical trials. One investigator may recruit subjects from several hospitals; one investigator may represent a team of clinicians (subinvestigators) who all recruit subjects from their own clinics at one hospital or at several associated hospitals. In case of any doubts regarding the definition of a clinical center in a statistical model, the statistical section of a clinical trial protocol (Subsection 1 of Section VI of these Guidelines) should clearly define the term (for example, the definition of a clinical center by investigator, location or region) in the context of a particular trial.

66. In most cases, centers can be satisfactorily determined through investigators at these centers. In case of any doubts, the aim should be to define the centers in such a way as to achieve uniformity with respect to important factors affecting the measurement of primary variables and the effect of treatments. Any rules for combining centers during the analysis should be justified in advance and set out, where possible, in the protocol of the clinical trial, but, in any case, decisions concerning this approach should always be taken blind to treatment, for example, at the time of the blind review.

67. The statistical model that will be used to evaluate and test treatment effects should be described in the protocol of the clinical trial. The main treatment effect may be studied first using a model that allows for center differences, but does not include a factor for «treatment-center interaction». If the treatment effect occurs homogeneously across centers, the routine inclusion of interaction factors in the model reduces the efficiency of the test

for the main effects. In the presence of true heterogeneity of treatment effects, the interpretation of the main treatment effect is controversial.

68. In some trials (for example, in some large mortality trials and participation of very few subjects per clinical center), there may be no reason to expect the influence of any center on the primary or secondary variables because they are unlikely to have clinically significant effect on the overall trial conclusion. In other clinical trials, it may be known from the start that the limited numbers of subjects per center will limit the value to include the center effects in the statistical model. In such cases, it is not appropriate to include the “center” term in the analysis model, and it is not necessary to stratify the randomization by center in this situation.

69. If positive treatment effects are found in a clinical trial with significant numbers of subjects per center, it is necessary, as a rule, to study the heterogeneity of treatment effects across centers, since it may affect the generalizability of the findings. Expressed heterogeneity may be identified by graphical presenting the results of individual centers or by using analytical statistical methods, such as a calculating the significance criterion for the «treatment-by-center» interaction. When using such a statistical significance criterion, it is important to understand that it has low power in a trial designed to detect the main effects of treatment.

70. If heterogeneity of treatment effects is detected, it should be interpreted with care and the professionals should actively search for explanation for this fact in terms of other features of trial management or subject characteristics. Such an explanation usually suggest appropriate further analysis and interpretation. To ensure the robustness of the estimates of treatment effect in the presence of heterogeneity of this treatment effect (which can be indicated by a expressed quantitative interaction between the treatment effect and the type of the clinical center), alternative estimates of

the treatment effect should be used with assignment of weighting coefficients corresponding to these centers to such alternative estimates from different clinical centers. Other approaches in this case are possible only if there is an appropriate rationale for their use in the clinical trial report. It is necessary to establish the cause of any heterogeneity of the treatment effect, characterized by expressed qualitative interactions between the treatment effect and the type of clinical center. If the reason for the treatment effect heterogeneity has not been established, additional clinical trials should be conducted until the treatment effect can be predicted with a sufficient degree of reliability.

71. The design of multicenter trials described in paragraphs 65-70 of these Guidelines is consistent with a statistical model with fixed effects. However, mixed models may also be used to study the heterogeneity of the treatment effect. These models consider the effects of «center» and «treatment-by-center» interaction as random factors and are most appropriate to clinical trials involving a large number of centers.

3. Type of Comparison

Superiority Trials

72. Scientifically, efficiency is most convincingly established by demonstrating:

superiority to placebo in a placebo-controlled trial;

superiority to an active control treatment in an active-control trial;

by establishing a «dose-response» relationship.

73. In the case of serious illnesses, if there is a therapeutic treatment, the efficiency of which is shown by superiority trial, a placebo-controlled trial may be considered unethical. In this case, the scientifically grounded use of an active treatment as a control should be considered. The use of placebo

control or active control should be rationalized in the protocol of each specific clinical trial.

Equivalence Trials and Non-Inferiority Trials of Medicinal Product

74. In some cases, an investigational medicinal product is compared to a reference medicinal product (treatment) without the objective of showing superiority of the investigational medicinal product. This type of trial is divided into two major categories according to its objective: the first is an equivalence trial, and the other is a non-inferiority trial.

75. Bioequivalence trials of generic medicinal products falls into the first category - that is, it refers to equivalence trials. In some cases, clinical equivalence trials are conducted for other reasons (for example, demonstrating the equivalence of a new investigational product to a reference medicinal product that is already marketed if these medicinal products are not absorbed when orally administered and do not enter the blood stream).

76. Many active control trials are designed to confirm that efficiency of an investigational product is not less than that of the active control, and therefore fall into the second category. Another option is a trial in which multiple doses of the investigational medicinal product are compared with a recommended dose or multiple doses of the medicinal product used for standard therapy. The purpose of this design is simultaneously to show a «dose-response» relationship for the investigational product and to compare the investigational product with the active control.

77. Equivalence and non-inferiority trials with active control may also include the additional use of placebo, thus pursuing multiple goals in one trial; for example, they may establish superiority to placebo, and thereby validate the trial design and simultaneously evaluate the degree of similarity

of efficiency and safety to the active control. The difficulties associated with the use of equivalence trials (and non-inferiority trials) with active control, which do not include a placebo or multiple doses of a new medicinal product, are well known. They arise from the inevitable lack of any measure of internal validity (in contrast to superiority trials), which makes external validation necessary.

78. Equivalence trials (and non-inferiority trials) is not conservative in nature, therefore, many flaws in the design and conduct of the trial will tend to bias the results in favor of the conclusion about equivalence. For these reasons, the design features of such trials should receive special attention and their conduct needs special care. For example, it is especially important to minimize the incidence of violations of the inclusion (non-inclusion) criteria, noncompliance, withdrawals, losses to follow-up, missing data, and other deviations from the protocol, and also to minimize their impact on the subsequent analyses.

79. Active comparators should be chosen with care. An example of a suitable active comparator would be a widely used therapy whose efficiency with appropriate indication was clearly established and quantified in well-designed and well-documented superiority trial(s) and which can be reliably expected to show similar efficiency in the planned active control trial. To this end, the new trial should have the same important design features (primary variables, the dose of the active comparator, eligibility criteria, etc.) as the previously conducted superiority trials in which the active comparator clearly demonstrated clinically significant efficiency, taking into account advances in medical or statistical practice relevant to the new trial.

80. It is of paramount importance that the protocol of a clinical trial designed to demonstrate the equivalence or non-inferiority contain a clear statement that the above constitutes its explicit objective. An equivalence

margin should be specified in the trial protocol as the largest difference in the efficiency evaluations, which can be considered clinically acceptable and should be smaller than differences in the efficiency evaluations demonstrated in superiority trials of the active comparator. In the case of the actively controlled equivalence trial, both the upper and lower margins are necessary, while only the lower margin is needed in actively controlled non-inferiority trial. The choice of equivalence margins should be rationalized clinically.

81. Statistical analysis is, as a rule, based on the use of confidence intervals. For equivalence trials, two-sided confidence intervals should be used. The conclusion about equivalence is made when the entire confidence interval falls within the equivalence margins. Technically, this is equivalent to the method of using two simultaneous one-sided tests to verify the (composite) null hypothesis that the treatment difference is outside the equivalence margins versus the (composite) alternative hypothesis that the treatment difference is within the margins. As the two null hypotheses are independent, the Type I error is properly controlled. For efficiency non-inferiority trials, a one-sided interval should be used. The approach based on the confidence intervals has an analogue with a one-sided hypothesis verification for testing the null hypothesis that the treatment difference (investigational product minus control) is equal to the lower margin of equivalence versus the alternative that the treatment difference is greater than the lower margin of equivalence. The choice of Type I error should be considered separate from the use of a one-sided or two-sided procedure. Sample size calculations should be based on these specified methods.

82. The conclusion about the equivalence or non-inferiority, based on obtaining an insignificant test result of the null hypothesis that there is no difference between the investigational medicinal product and the active comparator, is not acceptable.

83. There are also special issues in the choice of analysis sets. Subjects who are withdrawn or dropped out of the treatment group or the comparator group are more likely to have no response; thus, the results of using the full analysis set may introduce bias in favor of confirming equivalence of the investigational medicinal product.

Trials to Demonstrate «Dose-Response» Relationship

84. It is possible to study the correlation between the body's response to a new investigational medicinal product and its dose at all development phases and using any approaches in accordance with the Guidelines for the Selection of Doses of Medicinal Products (Annex to Recommendation No. 8 of the Eurasian Economic Commission's Board dated March 12, 2019). «Dose-response» trials may serve a number of objectives, including the following are of particular importance:

the confirmation of efficiency;

the investigation of the shape and location of the «dose-response» curve;

the estimation of an optimal starting dose;

the identification of optimal strategies for dose adjustment for individual needs;

the determination of a maximal dose beyond which additional benefit would be unlikely to occur.

85. These objectives should be considered using the data collected for a number of doses under investigation, including a placebo (zero dose) wherever appropriate. For this purpose, the application of procedures to estimate the relationship between dose and response, including the construction of confidence intervals and the use of graphical methods, is as

important as the use of statistical criteria. The methods used to verify hypotheses should be adapted taking into account the natural order of prescribing doses or specific questions regarding the shape of the «dose-response» curve (for example, monotonicity). The details of the planned statistical procedures should be given in the trial protocol.

4. Group Sequential Designs

86. Group sequential designs are used to facilitate the conduct of interim analysis. Despite the fact that acceptable types of designs permitting interim analysis are not limited to group sequential designs, they are most commonly used, since it is more practicable to evaluate the grouped outcomes of subjects at certain intervals during the trial, rather than on a continuous basis as data are received from each subject. The statistical methods should be fully specified before obtaining of information on treatment outcomes and subject treatment assignments (that is, before blind breaking). An independent data monitoring committee (Subsection 6 of Section V of these Guidelines) may be used to review or conduct the interim analysis of the data coming in as part of a group sequential design. Despite the fact that this design was most widely and most successfully used in large long-term trials of mortality or significant endpoints not related to mortality, its use is growing in other circumstances. In particular, there is an understanding that safety must be monitored in all trials, and therefore, the need for formal procedures determining early stopping for safety reasons should always be considered.

5. Sample Size

87. The number of subjects in a clinical trial should always be large enough to get a reliable answer to the questions posed. This number is

usually determined by the main objective of the trial. If the sample size is determined on some other basis, then this should be specified and rationalized. For example, a trial which sample size is calculated on the basis of safety questions or requirements or important secondary objectives may require larger numbers of subjects than a trial which size is calculated on the basis of the primary efficiency question, as indicated in the Guidelines on General Considerations for Clinical Trials (Annex to Recommendation No. 11 of the Eurasian Economic Commission's Board dated July 17, 2018).

88. Using the usual method for determining the appropriate sample size, the following items should be characterized: a primary variable, the statistical criterion, the null hypothesis, the alternative (working) hypothesis for the chosen dose(s) (including consideration of the treatment difference to be detected or rejected for the dose and in the subject population selected), the probability of erroneously rejecting the null hypothesis (the Type I error) and the probability of erroneously failing to reject the null hypothesis (the Type II error), as well as the approach to dealing with treatment withdrawals and protocol violations. In some cases, the event rate is of primary interest for evaluating power, and it is necessary to put forward assumptions to extrapolate from the required number of events to the final sample size for the trial.

89. The method for calculating the sample size should be given in the clinical trial protocol, along with estimates of all quantitative characteristics used in the calculations (such as variances, mean values, response rates, event rates, difference to be detected). The basis of these estimates should also be given. It is important to investigate the sensitivity of the sample size estimate to a variety of deviations from these assumptions; and this may be arranged by providing a range of sample sizes appropriate for a reasonable range of deviations from assumptions. In confirmatory trials, assumptions should, as a

rule, be based on published data or on the results of earlier trials. The treatment difference to be detected may be based on the assumption of the minimal effect which has clinical significance in the management of patients or on the assumption of the expected effect of the new treatment, if it is supposed to be larger than that of a standard treatment. Conventionally, the value of the probability of Type I error is taken to be 5 % or less or its value is established taking into account any adjustments made for reasons of multiplicity of comparisons; the specific choice may be influenced by the prior plausibility of the hypothesis under test and the desired impact of the results. The probability of Type II error is conventionally is taken to be 10 or 20 %; it is in the sponsor's interest to keep this figure as low as feasible, especially in the case of trials that are difficult or impossible to repeat. In some cases, acceptable or even preferred values are alternatives to generally accepted levels of Type I and Type II error.

90. Sample size calculations should provide the number of subjects required for the primary analysis. If this is the «full analysis set», estimates of the effect size may need to be reduced compared to the «per protocol set». This has to be to consider the “dilution” of the treatment effect arising due to the inclusion of data from patients who have withdrawn from treatment or who had poor compliance. The variability assumptions may also require revision.

91. The sample size of an equivalence trial or a non-inferiority trial should, as a rule, be based on the need to obtain a confidence interval for the treatment difference that shows that the treatments differ at most by a clinically acceptable difference. If the power of an equivalence trial is assessed from the assumption of a zero difference, then the sample size needed to achieve that power will be underestimated if the true difference is not zero. If the power of a non-inferiority trial is assessed for a true difference

of zero, then the sample size needed to achieve that power will be underestimated if the effect of the investigational product is less than that of the active control. The choice of a “clinically acceptable” difference requires rationale with respect to its meaning for future patients and may be less than the “clinically significant” difference mentioned above in the context of superiority trials designed to establish that a difference exists.

92. The exact sample size in a group sequential trial cannot be fixed in advance because it depends on randomness in combination with the chosen stopping guidelines and the true treatment difference. The design of the stopping guideline should take into account the consequent distribution of the sample size, usually specified as the expected and maximum sample sizes.

93. When event rates are lower than expected or variability is larger than expected, methods for sample size reestimation are available without unblinding data or making treatment comparisons.

6. Data collection and processing

94. The collection and transfer of data from the investigator to the sponsor can take place using a variety of media, including paper case record forms, remote site monitoring systems, medical computer systems, and electronic transfer. Regardless of the technology used to collect data, the form and contents of the information collected should be fully consistent with the protocol and should be established before the start of the clinical trial. The main focus should be on the data necessary to implement the planned analysis, including contextual information (for example, the timing assessments relative to dosage), required to confirm adherence to the protocol or to identify important protocol deviations. «Missing values» should be distinguishable from the «zero value» and «characteristic absence».

95. The process of data collection up to the database finalization should be carried out in accordance with Section 5 of the Good Clinical Practice Rules of the Eurasian Economic Union. In particular, timely and reliable processes for recording data and correcting errors and omissions are required to ensure that a quality database is obtained and trial objectives are achieved using the planned analysis.

V. Issues of Trial Conduct

1. Trial Monitoring and Interim Analysis

96. Careful high-quality conduct of a clinical trial according to the clinical trial protocol has a major impact on the reliability of the results. Careful monitoring can ensure that difficulties are detected early and their occurrence or recurrence minimized.

97. There are two different types of monitoring that generally characterize confirmatory clinical trials sponsored by the pharmaceutical industry. The first type of monitoring concerns the supervision of the quality of the trial, while the other type involves breaking the blind to make treatment comparisons (i.e. interim analysis). Both types of trial monitoring, in addition to different staff responsibilities, involve access to different types of trial data and information, and therefore, different principles apply for the control of potential statistical and operational bias.

98. In order to control the quality of the trial, the checks conducted to monitor the trial may include evaluating protocol compliance, acceptability of accumulated data, success of recruitment targets, relevance of design assumptions, success in keeping patients in the trials, etc. This type of monitoring does not require access to information on comparative treatment effects, unblinding of data and, therefore, has no impact on Type I error. The monitoring of a trial for this purpose is the responsibility of the sponsor and

can be carried out by the sponsor or an independent group selected by the sponsor. The period for monitoring this type usually begins with the selection of trial centers and ends with the collection and cleaning of data from the last subject of the trial.

99. The other type of clinical trial monitoring (interim analysis) involves the accumulation of comparative results of treatments. Interim analysis requires unblinded (that is, code breaking) access to treatment group assignment (actual treatment assignment or identification of group assignment) and comparative summary information on treatment groups. This requires that the protocol (or corresponding amendments prior to a first analysis) contain statistical plans for the interim analysis to prevent certain types of bias.

2. Changes in Inclusion and Non-inclusion criteria

100. Inclusion and non-inclusion criteria should remain constant (as specified in the trial protocol) throughout the entire period of subject recruitment. In some cases, changes may be appropriate (for example, in long-term trials, when new medical knowledge not related to direct trial data or based on the results of interim analyses suggests the need to change entry criteria). Changes may also be caused from the detection by monitors that regular violations of the inclusion (non-inclusion) criteria are occurring, or if, due to the over-restrictive criteria, there is a low rate of subject recruitment in the trial. Changes should be made without breaking the blind and always be described in a trial protocol amendment, which should cover all statistical consequences, such as sample size adjustments due to a different event rates, or modifying the planned analysis, such as stratifying the analysis in accordance with the modified inclusion (non-inclusion) criteria.

3. Accrual Rates

101. In trials with a long duration for the accrual of subjects, the rate of accrual should be monitored, and if the accrual rate falls significantly below the planned level, the reasons of dropping should be identified and remedial actions taken to protect the power of the trial and reduce concerns about selective inclusion in the trial and other aspects of quality. In multicenter trials, these measures are applied to individual centers as part of an entire clinical trial.

4. Sample Size Adjustment

102. In long-term trials, there will usually be an opportunity to check the assumptions, which underlie the original design and sample size calculations. This may be especially important if the trial features have been defined based on preliminary and (or) uncertain information. An interim check conducted using the blinded data may reveal that overall response variances, event rates or survival experience are not as expected. It is allowed, using appropriately modified assumptions, to recalculate the sample size, which should be rationalized and documented in a protocol amendment and in the clinical trial report. The measures taken to maintain blindness and the consequences (if any) for the value of the Type I error and the width of confidence intervals should be explained. The potential need to revise the sample size should be foreseen in the trial protocol whenever possible.

5. Interim Analysis and Early Stopping of Clinical Trial

103. An interim analysis is an analysis intended to compare treatment groups with respect to efficiency or safety at any time prior to formal completion of a trial. Since the number, methods, and consequences of these

comparisons affect the interpretation of the trial, all interim analyses should be carefully planned in advance and described in the trial protocol. Special circumstances may require an interim analysis that was not specified at the start of a trial. In these cases, a protocol amendment describing the interim analysis should be implemented prior to unblinded access to treatment comparison data. When an interim analysis is planned to decide whether or not to terminate a trial, a group sequential design is usually used, in which statistical monitoring schemes are used as rules.

104. The aim of such an interim analysis is to stop the trial early if the superiority of the studied treatment is clearly established; if the demonstration of a significant treatment difference has become unlikely, or if unacceptable adverse reactions have occurred. In general, the criteria (boundaries) for early termination of the trial for reasons related to efficiency require more evidence (that is, they are more conservative) than the criteria (boundaries) for monitoring safety. When the trial design and monitoring purpose involve multiple endpoints, then this aspect of multiple comparisons should also be considered when planning statistical data processing.

105. The trial protocol should describe the schedule for conducting interim analyses, or at least the rationale that will govern its generation (for example, whether flexible alpha spending function approaches will be used employed). Further details may be presented in a trial protocol amendment before the start of the first interim analysis. The stopping rules and their properties should be clearly described in the trial protocol or amendments. The potential effects of early trial stopping on the analysis of other important variables should also be considered. All these details should be written or approved by the Data Monitoring Committee, where applicable.

106. Deviations from the planned procedure always bear the risk that the results of the trial will not be valid. If it becomes necessary to make

changes to the trial, all subsequent changes to the statistical procedures should be described in an amendment to the protocol at the earliest possible opportunity, especially in terms of discussing the impact on the analysis and conclusions that such changes may cause. The procedures selected should always ensure that the overall probability of Type I error is controlled.

107. The execution of an interim analysis should be a completely confidential process because unblinded data and trial results are potentially affected. All staff involved in the conduct of the trial should remain blind to the results of such analyses, because of the possibility that their attitudes to the trial will be modified and cause changes in the characteristics of patients to be recruited or biases in treatment comparisons. This principle may be applied to all investigator staff and to staff employed by the sponsor except for those who are directly involved in the execution of the interim analysis. Investigators should be informed only about the decision to continue or to discontinue the trial, or to implement modifications to trial procedures.

108. Most clinical trials intended to support the efficiency and safety of an investigational product should continue until the completion of the recruitment of the entire planned sample size; trials should be stopped early only for ethical reasons or if the trial power is no longer considered acceptable. However, medicinal product development plans require the sponsor to have access to comparative treatment data for various reasons (for example, planning other trials). Portion of trials will include studying serious life-threatening outcomes or mortality, which, for ethical reasons, may require sequential monitoring of the accumulating comparative treatment data. In any of these situations, plans for interim statistical analysis should be in place in the trial protocol or in protocol amendments prior to the unblinded access to comparative treatment data in order to describe the method how to process potentially introduced statistical and operational bias.

109. In many clinical trials of investigational products, especially those that have major impact on public health, the responsibility for monitoring comparisons of efficiency and (or) safety outcomes should be assigned to an external independent group (independent data monitoring committee), a data and safety monitoring board, or a data monitoring committee, whose responsibilities should be clearly described.

110. If a sponsor assumes the role of monitoring efficiency or safety comparisons and therefore has access to unblinded comparative information, special care should be taken to protect the integrity of the trial and to appropriate management and limitation of the information sharing. The sponsor should ensure and document that the internal monitoring committee acted in accordance with written standard operating procedures and that minutes of decision-making meetings, including records of interim results, are maintained.

111. Any improperly planned interim analysis (with or without the further stopping the trial early) may flaw the results of a trial and possibly weaken reliability of the conclusions made. Therefore, such analyses should be avoided.

112. If unplanned interim analysis is conducted, the clinical trial report should explain why it was necessary and the degree to which blindness had to be broken, and provide an assessment of the potential magnitude of bias introduced and the impact of unplanned analysis on the interpretation of the results of the entire trial.

6. Role of Independent Data Monitoring Committee when assessing the progress of a clinical trial

113. An Independent Data Monitoring Committee may be established by the sponsor to periodically evaluate the progress of the clinical trial, safety

data, and critical efficiency variables and recommend to the sponsor whether to continue, modify or terminate a trial.

114. The Independent Data Monitoring Committee should have written operating procedures and maintain records of all its meetings, including interim results. Records should be made available to persons who monitor, audit, and inspect clinical trials upon completion of the clinical trial. The independence of this committee is intended to control the sharing of important comparative information and to protect the integrity of the clinical trial from adverse impact resulting from access to trial information. The Independent Data Monitoring Committee operates independently from an Institutional Review Board and an Independent Ethics Committee.

The Independent Data Monitoring Committee should include clinical trial scientists having knowledge in the appropriate disciplines, including statistics.

115. If the sponsor's representatives are part of the Independent Data Monitoring Committee, their rights and functions should be clearly defined in the Committee's operating procedures (for example, voting rights on key issues). Since these sponsor's representatives will have access to unblinded information, the procedures should also explain the control of sharing of interim trial results within the sponsor organization.

VI. Data Analysis

1. Preliminary Data Analysis Requirements

116. When designing a clinical trial, the principal features of the subsequent statistical analysis of the data should be described in the statistical section of the trial protocol. This section should include all the principal features of the proposed confirmatory analysis of the primary variable(s) and the way how to deal with the problems expected during the analysis. In the

case of exploratory trials, this section could include more general principles and directions.

117. The statistical analysis plan may be written as a separate document drawn up after the final approval of the trial protocol. In terms of statistical analysis, a more technical and detailed elaboration of the main characteristics stated in the protocol may be included. The plan of statistical analysis may include detailed procedures for executing the statistical analysis of the primary and secondary variables and other data. The plan of statistical analysis should be reviewed and updated as a result of the blind review of the data (where necessary) and should be finalized before breaking the blind. It is necessary to keep formal documents about the time when the statistical analysis plan was finalized as well as when the blind was subsequently broken.

118. If the blind review suggests changes to the principal features stated in the trial protocol, it should be documented in a protocol amendment. Otherwise, it would be enough to update the statistical analysis plan in accordance with the recommendations of the blind review. Only results from analyses envisaged in the trial protocol (including amendments) may be regarded as confirmatory.

119. In the statistical section of the clinical trial report, the statistical methodology should be clearly described including when in the clinical trial process methodology decisions were made.

2. Analysis Sets

120. The set of trial subjects whose data are to be included in the main analyses should be defined in the statistical section of the trial protocol. Data documentation for all subjects for whom trial procedures were initiated (for example, run-in period) should be envisaged. The content of the documented

data for the subject depends on detailed features of the particular trial, but at least demographic and baseline data on disease status should be collected whenever possible.

121. If all subjects randomized into a clinical trial satisfied all inclusion (non-inclusion) criteria, followed all trial procedures perfectly with no losses to follow-up, and provided complete data records, then the set of subjects to be included in the analysis would be self-evident. The design and conduct of a trial should be aimed at maximum compliance with such an ideal, but in practice, it is doubtful if it can be fully achieved. Therefore, the statistical section of the trial protocol should reflect in advance the anticipated problems in terms of how they will affect the subjects and data to be analyzed. The trial protocol should also establish procedures aimed at minimizing all anticipated errors in trial conduct that might impair a satisfactory analysis, including various types of trial protocol violations, presence of withdrawals and missing values. The trial protocol should also provide ways to reduce the frequency of such problems and methods for solving problems that arise during data analysis. Possible amendments to the method of handling protocol violations during data analysis should be identified during the blind review. It is desirable to identify time when every important violation of trial protocol occurred, its causes, and its influence on the trial result. The frequency and type of trial protocol violations, missing values, and other problems should be documented in the clinical trial report and their potential influence on the trial results should be described.

122. Decisions concerning the analysis set should be based on the following principles:

a) to minimize bias; and

b) to avoid inflation of Type I error.

Full Analysis Set

123. The «intention-to-treat» principle implies that the primary analysis should include all randomized patients. Compliance with this principle would require complete follow-up of all randomized subjects for the studied outcomes. In practice, this ideal may be difficult to achieve due to the reasons described below. In these Guidelines, the «full analysis set» concept is used to describe the analysis set that most completely and most closely matches the «intention-to-treat» concept including all randomized subjects.

124. Preservation of the initial randomization in analysis is important to prevent bias and to provide a secure foundation for statistical tests. In many clinical trials, the use of the full analysis set provides a conservative strategy. In many situations, it may also provide estimates of treatment effects that are more likely to reflect those observed in subsequent practice.

125. There are a limited number of circumstances that might lead to excluding randomized subjects from the full analysis set, including the failure to satisfy principal inclusion (non-inclusion) criteria (eligibility violations), cases where a subject fails to take a single dose of investigational medicinal product, and the lack of any data after randomization. Such exclusions should always be rationalized. Subjects who fail to satisfy an inclusion (non-inclusion) criterion may be excluded from the analysis without the possibility of introducing bias only under the following circumstances:

a) the entry criterion was measured prior to randomization;

b) the detection of the relevant eligibility violations can be made completely objectively;

c) all subjects are thoroughly assessed for eligibility violations in the same manner (this may sometimes be difficult to achieve in an open-label trial, or even in a double-blind trial, if the data are unblinded prior to this check, which emphasizes the importance of the blind review);

d) all detected violations of the some inclusion (non-inclusion) criterion are excluded.

126. In some situations, it may be reasonable to eliminate from the set of all randomized subjects any subject who took no investigational medicinal product. The «intention-to-treat» principle would be preserved despite the exclusion of these subjects provided, for example, that knowledge of the prescribed treatment could not affect the decision to initiate the treatment. In other situations, it may be necessary to eliminate from the set of all randomized subjects any subject who does not have any data after randomization.

No analysis should be considered complete unless the potential biases arising from these specific exclusions, or any others, are considered.

127. If the full analysis set of subjects is used, violations of the protocol that occur after randomization may have an impact on the data and conclusions, particularly if their occurrence is related to treatment assignment. In most cases, the inclusion of data from such subjects in the analysis is acceptable and complies with the «intention-to-treat» principle. Special problems arise in connection with subjects withdrawn from treatment after receiving one or more doses who do not provide data after the withdrawn point, and subjects who are lost to follow-up for other reasons, because failure to include these subjects in the full analysis set may seriously undermine the approach. In this context, measurements of primary variables made at the time of the loss of a subject to follow-up for any reason, or subsequently collected in accordance with the intended schedule of assessments in the protocol, are valuable; subsequent collection is especially important in trials where the primary variable is mortality or serious morbidity. The intention to collect data in this way should be described in the protocol. Imputation techniques (data completion), ranging of which is

widely represented from the Last Observation Carried Forward (LOCF) to complex mathematical models, may also be used to compensate for missing data.

128. Other methods employed to ensure the availability of measurements of primary variables for every subject in the full analysis set may require some assumptions about the subjects' outcomes or a simpler choice of outcome (for example, success – failure). The use of any of these strategies should be described and rationalized in the statistical section of the trial protocol, and the assumptions underlying any mathematical models employed should be clearly explained. It is also necessary to confirm the robustness of the corresponding results of analysis, especially when the strategy under consideration could itself lead to biased estimates of treatment effects.

129. Due to the unpredictability of some problems, it may sometimes be preferable to postpone detailed description of the solutions to these issues until the blind review of the data at the end of the trial, and, if such problems are expected, this approach should be reflected in the trial protocol in advance.

«Per Protocol» Set

130. The «per protocol» set of subjects, sometimes described as the «valid cases», the «efficiency» sample, or the «evaluable subjects» sample, defines a subset of the subjects in the full analysis set who are more compliant with the protocol and is characterized by the following criteria:

a) obtaining a predefined minimal exposure of prescribed treatment;

b) the availability of measurements of the primary variable(s);

c) the lack of any major protocol violations, including the violation of inclusion/non-inclusion criteria.

131. The precise reasons for excluding subjects from the «per protocol» set should be fully characterized and documented before breaking the blind according to the circumstances of the specific trial.

132. The use of the «per protocol» set may maximize the opportunity for a new treatment to confirm additional efficiency in the analysis, and most precisely reflects the scientific model underlying the protocol. However, the corresponding hypothesis verification and assessment of the treatment effect may or may not be conservative, depending on the trial; the bias, which may be severe, arises from the fact that adherence to the trial protocol may be related to treatment and outcome.

133. All problems that lead to the exclusion of subjects to generate the «per protocol» set, and other protocol violations, should be fully identified and summarized. This group of violations of the trial protocol includes errors in treatment assignment, the use of a medicinal product prohibited by the trial protocol, poor compliance, and loss for follow-up and missing data. It is good practice to assess the pattern of such problem occurrence in the treatment groups in terms of frequency and time before they occur.

Roles of the Different Analysis Sets

134. In general, it is advisable to demonstrate the lack of sensitivity of the principal trial results to alternative choices of the set of subjects analyzed. In confirmatory trials, it is usually necessary to plan to conduct both an analysis of the full analysis set and a «per protocol» analysis, so that any differences between them can be the subject of open discussion and interpretation. In some cases, it may be desirable to plan a more profound study of the sensitivity of conclusions to choosing the set of subjects analyzed. When the full analysis set and the «per protocol» set lead to fundamentally the same conclusions, confidence in the trial results is

increased, bearing in mind, however, that the need to exclude a substantial proportion of subjects from the «per protocol» analysis casts some doubt on the overall validity of the trial.

135. The full analysis set and the «per protocol» set play different roles in superiority trials (which aimed at showing the investigational product to be superior) and in equivalence or non-inferiority trials (which aimed at showing the investigational product to be comparable). In superiority trials, the full analysis set is used in the primary analysis (with rare exception) because it allows avoiding over-optimistic estimates of efficiency resulting from a «per protocol» analysis, as the noncompliant subjects included in the full analysis set tend, as a rule, to reduce the estimated treatment effect. Whereas in an equivalence or non-inferiority trial, use of the full analysis set is, as a rule, not a conservative approach and its role should be considered carefully.

3. Missing Values and Outliers

136. Missing values are a potential source of bias in a clinical trial. Therefore, it is necessary to ensure maximum compliance with all the requirements of the trial protocol in relation to the collection and handling of data. During any clinical trial, there are always some missing data (values). Nevertheless, a trial may be considered as valid, if the adequate methods of handling missing data (values) are used, particularly, if they are predefined in the trial protocol. Definition of methods may be optimized by updating this aspect in the statistical analysis plan during the blind review. Currently, there are no universal methods for handling missing data (values). It is necessary to study the sensitivity of the analysis results to the method of handling missing data (values), especially if the number of missing data (values) is substantial.

137. A similar approach should be used when assessing the influence of outliers, the statistical definition of which is, to some extent, arbitrary. Clear

attribution of a particular value to an outlier is most convincing when rationalized both medically and statistically, then the medical context will then often define the appropriate action. No outlier procedure set out in the protocol or the statistical analysis plan should support any treatment group a priori. It is worth reiterating that this aspect of the analysis can be updated during blind review which would favorably affect the quality of the analysis. If the trial protocol does not provide for a procedure for dealing with outliers, it is necessary to conduct one analysis with all the values obtained and at least one analysis after eliminating or reducing the outlier effect and discuss the difference between their results.

4. Data Transformation

138. The decision to transform key variables prior to analysis is best made at the design stage of the trial on the basis of similar data from earlier clinical trials. Transformations (for example, square rooting, logarithm) should be established in the protocol and provide a rationale, especially for the primary variable(s). The general principles guiding the use of transformations aimed at ensuring that the assumptions underlying the statistical methods are met should be given in standard sections; in some clinical areas, standard approaches to specific variables have been developed. The decision on whether and how to transform a variable should depend on the choice of a specific scale that facilitates clinical interpretation.

139. Similar approaches are applicable to other derived variables, such as the change from baseline, percentage change from baseline, the «area under the curve» of repeated measures, and the ratio of two different variables. Subsequent clinical interpretation of results should be carefully considered, and the use of derived variables should be rationalized in the trial protocol.

5. Estimation, Confidence Intervals, and Hypothesis Verification

140. The statistical section of the trial protocol should include the hypotheses that are to be verified and (or) the treatment effects that are to be estimated in order to satisfy the primary objectives of the trial. The statistical methods used to accomplish these tasks should be described for the primary (and preferably the secondary) variables, and the underlying statistical model should be explained. In all possible cases, estimates of treatment effects should be accompanied by confidence intervals specifying how to calculate them. A description should be given of any intentions to use baseline data to improve accuracy (precision) or to adjust estimates for potential baseline differences, for example, by means of analysis of covariance.

141. It is important to specify which tests for statistical significance will be used: whether one- or two-sided tests. The use of one-sided tests should be rationalized in advance. If hypothesis verification is not considered appropriate, then the alternative process for making statistical conclusions should be given. The issue of using one-sided or two-sided approaches to draw conclusions is controversial, and a diversity of views can be found in the statistical literature. The approach of setting Type I errors for one-sided tests equal to half the standard Type I error for two-sided tests is preferable in regulatory conditions. This enhances consistency with the two-sided confidence intervals that are, as a rule, appropriate for estimating the possible size of the difference between two treatments.

142. The particular statistical model chosen should reflect the current state of medical and statistical knowledge about the analyzed variables as well as the statistical design of the trial. All effects included in the analysis (for example, in analysis of variance models) should be fully characterized, and the method, if any, by which this set of effects might be modified in

response to preliminary results should be explained. The same considerations apply to the set of covariates included in an analysis of covariance. In the choice of statistical methods, due attention should be paid to the statistical distribution of both primary and secondary variables. When making this choice (for example, between parametric and nonparametric methods), it is important to bear in mind the need to provide statistical estimates of the size of treatment effects together with confidence intervals (in addition to significance tests).

143. The primary analysis of the primary variable should be clearly distinguished from supporting analyses of the primary or secondary variables. In the statistical section of the trial protocol or statistical analysis plan, the information on the method of generalization and registration of data other than the primary and secondary variables should be included. Such information should include a description of all approaches adopted to achieve consistency of analysis between different trials (for example, regarding safety data of a medicinal product).

144. Modeling approaches that consider information on known pharmacological parameters, the extent of protocol compliance for individual subjects, or other biologically based data may provide valuable insights into actual or potential efficiency, especially with regard to estimation of treatment effects. The assumptions underlying such models should always be clearly identified, and the limitations of any conclusions should be carefully described.

6. Adjusting Significance Levels and Confidence Interval

145. In the case of multiplicity, the usual frequentist approach to the analysis of clinical trial data may require an adjustment to the Type I error.

Multiplicity may arise, for example, from multiple primary variables, multiple comparisons of treatments, repeated evaluation over time, and (or) interim analyses. Sometimes, when available, methods of avoiding or reducing multiplicity are preferable, such as establishing a key primary variable (multiple variables), choosing the critical treatment contrast (multiple comparisons), and using a summary parameter such as “area under the curve” (repeated measurements). In confirmatory analyses, any aspects of multiplicity that persist after steps of this kind have been taken should be specified in the protocol; the analysis plan should always consider making adjustments and describe the features of the adjustment procedure or explaining why it is not required.

7. Subgroups, Interactions, and Covariates

146. The primary variable(s) is often systematically related to other influences apart from treatment. For example, there may be a relationship to covariates such as age and sex, or there may be differences between specific subgroups of subjects, such as subjects treated at the different centers of a multicenter trial. In some instances, an adjustment for the influence of covariates or for subgroup effects is an integral part of the planned analysis and therefore should be set out in the protocol. When planning a trial, it is necessary to identify in advance those covariates and factors expected to have an important effect on the primary variable(s), and to propose ways how to account for these in the analysis in order to improve accuracy (precision) and compensate for the lack of balance between treatment groups. If one or more factors are used to stratify the design, it is appropriate to account for those factors in the analysis. In case of doubt about the potential value of the adjustment, it is often recommended to nominate the unadjusted analysis as the one for primary attention, and the adjusted analysis as supportive one.

Special attention should be paid to center effects and to the role of baseline measurements of the primary variable. It is not advisable to introduce adjustments in the main analyses for covariates measured after randomization because they may be affected by the treatments.

147. The treatment effect itself may vary depending on subgroup or covariate: for example, the effect may decrease with age or may be larger in a particular diagnostic category of subjects. In some cases such interactions are anticipated or are of particular prior interest (for example, geriatrics); and hence a subgroup analysis or a statistical model accounting for interactions is part of the planned confirmatory analysis. In most cases, subgroup or interaction analyses are exploratory and should be clearly identified as such; they would help to evaluate the uniformity of all overall treatment effects found. In general, such analyses should be started by adding interaction terms to the statistical model in question, followed by additional exploratory analysis within relevant subgroups of subjects, or within strata defined by the covariates. If such analyses are conducted as exploratory, they require careful interpretation, any conclusion about the treatment efficiency (or lack thereof) or safety based solely on exploratory analyses in subgroups, is unlikely to be accepted.

8. Integrity of Data and Computer Software Validity

148. The reliability of the numerical results of the analysis depends on the quality and validity of the methods and software (both internally and externally written) used both for data management (data entry, storage, verification, correction, and retrieval) and for processing the data statistically. In this regard, data management activities should be based on thorough and effective standard operating procedures. The computer software used for data

management and statistical analysis should be reliable, and documentation of appropriate software testing procedures should be available.

VII. Evaluation of Safety and Tolerability

1. Scope of Safety and Tolerability Evaluation

149. In all clinical trials, evaluation of safety and tolerability constitutes an important element. In early phases, this evaluation is mostly of an exploratory nature and is only sensitive to frank expressions of toxicity, whereas in later phases the establishment of the safety and tolerability profile of a medicinal product can be characterized more fully in larger samples of subjects. Controlled trials in later phases are an important means of objective exploring of any new potential adverse reactions, even if such trials generally lack power in this respect.

150. Separate trials may be designed with the purpose of making specific hypotheses about superiority or equivalence with regard to safety and tolerability compared to another medicinal product or to another dose of the investigational medicinal product. Such specific hypotheses should be supported by relevant evidence from confirmatory clinical trials, similar to what is done for corresponding efficiency hypotheses.

2. Choice of Variables and Data Collection

151. In any clinical trial, the methods and parameters chosen to evaluate the safety and tolerability of a medicinal product depend on a number of factors, including knowledge of the adverse reactions of closely related medicinal products, information from nonclinical and earlier clinical trials and possible consequences of the pharmacodynamic and pharmacokinetic properties of the particular medicinal product, route of administration, the type of subjects to be studied, and the duration of the trial.

152. The main scope of data on the safety and tolerability of a medicinal product is usually formed by laboratory tests in terms of clinical biochemistry and clinical blood tests, vital signs and clinical adverse events (diseases, symptoms and complaints). The occurrence of serious adverse events and treatment discontinuations due to adverse events are particularly important to register.

153. Moreover, it is recommended that a consistent methodology be used for the data collection and evaluation throughout a clinical trial program to facilitate the consolidation of data from different trials. The use of a common adverse event dictionary is particularly important. This dictionary has a structure that makes it possible to summarize the adverse event data on three different levels: system-organ class, preferred term, or included term. The preferred term is the level on which adverse events usually are summarized; and preferred terms belonging to the same system-organ class could then be combined in the descriptive presentation of data.

3. Set of Subjects to Be Evaluated and Presentation of Data

154. For the final safety and tolerability assessment, the set of subjects to be summarized is usually defined as those subjects who received at least one dose of the investigational medicinal product. Safety and tolerability variables should be collected as comprehensively as possible from these subjects, including type of adverse event, severity, onset, and duration. Additional safety and tolerability evaluations may be required in specific subpopulations, such as females, the elderly, the severely ill, or those who have a common concomitant treatment. These evaluations may need to address more specific issues.

155. All safety and tolerability variables will need attention during evaluation, and the broad approach should be indicated in the trial protocol. All adverse events should be reported, whether or not they are considered to be related to treatment. All available data in the studied population should be accounted for in the evaluation. Definitions of measurement units and reference ranges of laboratory variables should be made with care; if different units or different reference ranges are used in the same trial (for example, if several laboratories are involved), then measurements should be appropriately standardized to allow a unified evaluation. Use of a toxicity grading scale should be envisaged in advance and rationalized.

156. The incidence of a certain adverse event is usually expressed as the ratio of the number of subjects experiencing events to the number of subjects at risk. However, it is not always apparent how to assess incidence. For example, depending on the situation, the number of exposed subjects or the extent of exposure (in person-years) could be used as the denominator. Whether the purpose of the calculation is to estimate a risk or to make a comparison between treatment groups, it is important that the definition be given in the trial protocol. This is especially important if long-term treatment is planned and a substantial proportion of treatment withdrawals or deaths are expected. For such situations, survival analysis methods should be considered and cumulative adverse event rates calculated in order to avoid the risk of underestimation.

157. In situations if there is a substantial background noise of symptoms and complaints (for example, in psychiatric trials), when assessing the risk of various adverse events, it is necessary to provide methods for its consideration. One of these methods is the use of the «treatment emergent» concept, according to which adverse events are recorded only if they occur or worsen compared to the pretreatment period.

158. Other methods to reduce background noise may also be appropriate, such as ignoring adverse events of mild severity or requiring that an event be persisted at repeated visits to be included as the numerator. Such methods should be described and rationalized in the trial protocol.

4. Statistical Evaluation

159. Studying safety and tolerability of a medicinal product is a comprehensive challenge. Although some specific adverse reactions can usually be foreseen and specifically monitored for any medicinal product, the range of possible adverse reactions is very large, and new and unforeseeable reactions are always possible. Moreover, an adverse event that occurs after a protocol violation, such as the use of a prohibited medicinal product, can introduce a bias. This background causes the statistical difficulties associated with the analytical evaluation of safety and tolerability of medicinal products, and means that unequivocal information from confirmatory clinical trials is the exception rather than the rule.

160. In most trials, the safety and tolerability effects are best-analyzed using descriptive statistical methods to the data, supplemented by calculation of confidence intervals if this helps interpretation. It is also valuable to use graphical methods for presenting data in which patterns of distribution of adverse events are given for both treatment groups and subjects.

161. In some cases, calculating p-values is useful either to help evaluate some difference of interest or as a «marker» applied to a large number of safety and tolerability variables to highlight differences worthy of further attention. This is particularly useful for laboratory data, which otherwise can be difficult to summarize appropriately. Laboratory data are recommended to be subjected to quantitative analysis (for example, evaluation of treatment

means) and to qualitative analysis (for example, by counting values above or below a certain threshold).

162. If hypothesis verifications are performed, statistical adjustments for multiplicity to quantify the Type I error are appropriate, but the Type II error is usually of more concern. Putative statistically significant results should be carefully interpreted when there is no multiplicity adjustment.

163. In the majority of trials, investigators try to establish that there are no clinically unacceptable differences in terms of safety and tolerability compared with either a reference medicinal product or a placebo. Similar to the efficiency evaluation in trials of equivalence and non-inferiority, the use of confidence intervals is preferred to hypothesis testing in this situation. In this way, a low degree of accuracy (precision) of the data, which often occurs due to the low frequency of occurrence, is clearly demonstrated.

5. Integrated Summary

164. The safety and tolerability properties of a medicinal product are usually summarized from various trials continuously during an investigational product's development and, in particular, at the time of submitting a marketing application. The usefulness of this summary is dependent on adequate and well-controlled trials with high data quality.

165. The overall usefulness of a medicinal product is always a question of balance between risk and benefit, and such a perspective could also be addressed in a single trial even if the assessment of «risk-benefit» ratio usually is performed upon the summary of the entire clinical trial program.

VIII. Reporting Results

1. Evaluation and Reporting Results

166. During the trial planning, the trial protocol should specify the principal features of the analysis described in Section VI of these Guidelines. When the conduct of the trial is over and the data are assembled and available for preliminary inspection, it is advisable to carry out the blind review of the planned analysis described in Section VI of these Guidelines. This preliminary review, blinded to treatment, should cover decisions concerning, for example, the exclusion of subjects or data from the analysis sets; the checking of possible transformations and detections of outliers are also permissible; important covariates found in recent trials may be added to the model; the decision to use parametric or nonparametric methods may be revised. Decisions made at this time should be described in the report and should be distinguished from the decisions made after the statistician has had access to the treatment codes, as blind decisions, as a rule, have less potential to introduce bias. Statisticians or other staff involved in unblinded interim analysis should not participate in the blind review or in making modifications to the statistical analysis plan. If the blinding is compromised due to the possibility that treatment-induced obvious effects may appear in the data, special care should be taken for the blind review.

167. Many of the more detailed aspects of presentation and tabulation should be finalized during or shortly before the blind review so that, by the time of the actual analysis, full plans are prepared for all its aspects including subject selection, data selection and modification, data summary and tabulation, estimation, and hypothesis verification. Once data validation is complete, the analysis should proceed according to the predefined plans; the more these plans are adhered to, the greater the reliability of the results. Special attention should be paid to any differences between the planned analysis as described in the protocol, the protocol amendments, or the statistical analysis plan updated on the basis of a blind review of data, and the

actual analysis. A careful explanation should be provided for deviations from the planned analysis.

168. All subjects who entered the trial should be accounted for in the trial report, whether or not they are included in the analysis. All reasons for exclusion from analysis should be documented; for any subject included in the full analysis set but not in the «per protocol» set, the reasons for subject exclusion from the «per protocol» set should also be documented. Similarly, for all subjects included in an analysis set, the values of all important variables should be accounted for at all significant time-points.

169. The effect of all losses of subjects or data, withdrawals from trials, and major protocol violations on the main analyses of the primary variable(s) should be evaluated carefully. Subjects lost to follow-up, withdrawn from treatment, or with a severe protocol violation should be identified and a descriptive analysis of them should be provided, including the reasons for their withdrawal or exclusion from the trial and the relationship of these reasons directly to the treatment and studied outcome.

170. Descriptive statistics form an integral part of reports. Relevant tables and (or) graphical presentations should illustrate clearly the important features of the primary and secondary variables and of key prognostic and demographic variables. The results of the main analyses relating to the objectives of the trial should be the subject of particularly careful descriptive presentation. When reporting the results of significance tests, precise p-values (for example, $p = 0.034$) should be reported rather than making exclusive reference to critical values.

171. The primary goal of the analysis of a clinical trial should be to answer the questions posed by its main objectives. Moreover, new questions based on the observed data and that require statistical evaluation may arise during the unblinded analysis. Additional statistical analysis or complication

of the already planned analysis may be the consequence. This additional work should be given in a separate section of the clinical trial report and separated from the work planned by the trial protocol.

172. There is the chance of unforeseen imbalances between the treatment groups in terms of baseline indicators that are not predefined as covariates in the planned analysis, but have some prognostic importance. The optimal solution to this situation is achieved by confirming that an additional analysis, taking into account these imbalances leads, in essence, to the same conclusions as the planned analysis. If this is not achieved, the effect of unforeseen imbalances on the findings should be discussed.

173. In general, unplanned data analyses should be used to a limited extent. Such analyses are often carried out when it is assumed that the treatment effect may vary according to some other factor or factors. An attempt may then be made to identify subgroups of subjects for whom the effect is particularly beneficial. The potential dangers of over-interpretation of unplanned subgroup analyses are well known (see Subsection 7 of Section VI of these Guidelines) and should be carefully avoided. Although similar problems of interpretation may arise if a treatment appears to be ineffective or have adverse reactions in the subgroup of subjects, such possibilities should be properly assessed and the appropriate presentation of the obtained data should be subsequently ensured.

174. A statistical conclusion should be included in the analysis, interpretation and presentation of the results of a clinical trial. To this end, the trial statistician should be a member of the team responsible for the clinical trial report and should approve the clinical report.

2. Summarizing the Clinical Database

175. An overall summary and synthesis of all evidence on safety and efficiency from all the reported clinical trials should be provided for a marketing application. This may be accompanied, when appropriate, by a statistical combination of results.

176. As part of this summary, a number of specific statistical problems arise: describing the demography and clinical features of the population treated during the course of the clinical trial program; addressing the key questions of efficiency by considering the results of the relevant (usually controlled) trials and highlighting the degree to which their results reinforce or contradict each other; summarizing the safety information available from the combined database of all the trials whose results contribute to the marketing application, and identifying potential safety issues. During the design of a clinical program, careful attention should be paid to the uniform definition and collection of measurements, which will facilitate further interpretation of the series of trials, particularly if measurements can be combined across trials. A common dictionary for recording the information about the medicinal product, medical history and adverse events should be selected and used. A common definition of the primary and secondary variables is nearly always required, and it is essential for meta-analysis. The way of measuring key efficiency variables, the timing of assessments relative to randomization and entry, the handling of protocol violators and deviators, and perhaps the definition of prognostic factors should all be kept consistent between individual trials unless there are valid reasons not to do so.

177. All statistical procedures used to combine data across trials should be described in detail. Special attention should be paid to the possibility of bias associated with the selection of trials, to the homogeneity of their results, and to the proper modelling of the various sources of variation. The

sensitivity of conclusions to the assumptions and selection options should be evaluated.

Efficiency Data

178. Individual clinical trials should always be large enough to achieve their objectives. Additional valuable information may also be gained by summarizing a series of clinical trials that address essentially identical key efficiency questions. The main results of such a set of trials should be presented in an identical form to make comparisons, usually in tables or graphs that focus on estimates plus confidence limits. The use of meta-analytic techniques to combine these estimates is often a useful addition because it allows generating a more precise overall estimate of the size of the treatment effects and provides a complete and concise summary of the results of the trials. In exceptional cases, a meta-analysis may also be the most appropriate way, or the only way, to provide sufficient overall evidence of efficiency by verifying a general hypothesis. When used for this purpose, the meta-analysis should have its own prospectively created protocol.

Safety Data

179. When summarizing safety data, it is important to examine the safety database for any indications of potential toxicity and to monitor any indications in the future using an appropriate supportive observation model. The combination of the safety data over all human use of (exposure to) the medicinal product is an important source of information because the larger sample size increases the chance of detecting the rarer adverse events and, perhaps, of estimating their approximate frequency. However, frequency data from this database are difficult to evaluate because of the lack of a comparator group, and data from comparative trials are especially valuable in

overcoming this difficulty. The results of trials that use the same comparator (placebo or specific active comparator) should be combined and presented separately for each comparator providing sufficient data.

180. All indications of potential toxicity arising from data studying should be reported. The evaluation of the reality of these potential adverse reactions should take into account the issue of multiplicity arising from the numerous comparisons made. Survival analysis methods in order to obtain a potential relationship between the frequency of adverse events and the duration of exposure and (or) follow-up should be used properly during the evaluation. The risks associated with the identified adverse reactions should be quantitatively assessed in an appropriate manner to ensure a proper evaluation of the risk-benefit ratio.