REVIEW ARTICLE

Dan L. Longo, M.D., *Editor*

STATISTICS IN MEDICINE

Multiplicity Considerations in Clinical Trials

Alex Dmitrienko, Ph.D., and Ralph B. D'Agostino, Sr., Ph.D.

From Mediana, Overland Park, KS (A.D.); and Boston University, Boston (R.B.D.). Address reprint requests to Dr. Dmitrienko at alex.dmitrienko@gmail.com.

N Engl J Med 2018;378:2115-22.
DOI: 10.1056/NEJMra1709701

MULTIPLICITY, OR THE USE OF MANY COMPARISONS IN A CLINICAL TRIAL, increases the likelihood that a chance association could be deemed causal. This problem commonly arises in clinical trials that have several clinical objectives based on the evaluation of multiple end points or multiple dose–control comparisons, evaluation of several patient populations, and other factors. Multiplicity considerations play a central role in the assessment of efficacy evidence in the presence of competing clinical objectives. The more comparisons that are made, the more likely it is that a comparison that appears to be significant will be falsely so.

The selection of an appropriate statistical strategy for dealing with multiplicity is critical for performing reliable inferences and maximizing the probability of success in a clinical trial. Early articles on multiplicity problems arising in clinical trials were published from the 1960s through the 1990s.[1-4] This topic has attracted much attention in the clinical trial literature, and numerous new statistical approaches to performing multiplicity adjustments have appeared since the 1990s.[5,6]

Regulatory agencies around the world have recognized the importance of addressing multiplicity in confirmatory phase 3 clinical trials. Requirements for well-controlled clinical trials to support a new indication include prespecification of appropriate statistical methods for controlling the type I error rate (the false positive rate, or the rate of falsely rejecting a true null hypothesis). The Food and Drug Administration (FDA) has recently released comprehensive guidance on handling multiple end points in clinical trials,[7] and the European Medicines Agency (EMA) has also published guidelines on multiplicity issues in clinical trials.[8,9]

In this review, we describe common multiplicity problems in clinical trials, as well as statistical methods aimed at achieving control of the error rate (known as multiplicity adjustments). We also use case studies to illustrate recommended approaches. We focus on confirmatory phase 3 trials that are conducted to pursue specific efficacy claims. However, proper treatment of multiplicity issues also plays an important role in exploratory trials (e.g., dose-finding phase 2 trials) aimed at hypothesis generation.

OVERVIEW OF MULTIPLICITY PROBLEMS IN CLINICAL TRIALS

As a working definition, multiplicity is defined here as simultaneous evaluation of multiple aspects of the efficacy profile of a given treatment.[10]

CLASSIFICATION OF MULTIPLICITY PROBLEMS

A wide variety of multiplicity problems may be encountered in clinical trials. When multiplicity is caused by a single factor (e.g., analysis of multiple end points), the problems are often referred to as problems with a single source of multiplicity. It

is increasingly common to formulate complex sets of clinical objectives in phase 3 trials. Examples are trials with multiple end points evaluated at several dose levels of an experimental treatment or in several patient populations. In such trials, there are several sources of multiplicity, leading to more complex problems.

It is important to understand the role of "win criteria," or clinical decision rules, in clinical trials.[5] Most commonly, the overall outcome of a trial is declared to be positive if at least one of the predefined clinical objectives is met. For example, a phase 3 trial with several primary end points is declared successful if a significant treatment effect is established on one or more of these end points. Performing these data analyses without a proper statistical adjustment leads to an inflated probability of incorrect conclusions. This probability of incorrect conclusions is known as the overall type I error rate (also known as the familywise error rate). In confirmatory clinical trials, control of the type I error rate at a two-sided 5% level is mandated by regulatory agencies[7-9] to enable formulation of specific efficacy or safety claims. We focus on multiplicity issues in phase 3 trials that use this criterion.

Clinical decision rules can also be defined on the basis of a simultaneous analysis of the end points of interest, known as coprimary end points.[11,12] A successful outcome can be claimed in a trial if all clinical objectives are met. In clinical trials of an experimental drug versus placebo for the treatment of Alzheimer's disease, a significant improvement in the coprimary end points, the ADAS-cog (Alzheimer's Disease Assessment Scale–cognitive subscale) score and the CIBIC-Plus (Clinician's Interview-Based Impression of Change Plus Caregiver Input) score, is required to support an effectiveness claim. Since a beneficial treatment effect on all end points must be established, the type I error rate is not inflated in clinical trials with coprimary end points, and no formal multiplicity adjustment is applied. Rather, the treatment effect on each end point is tested at a standard 5% level, which results in a conservative approach, since the actual type I error rate is likely to be much lower than 5%.

MULTIPLICITY ISSUES IN EXPLORATORY TRIALS

Statistical methods to control the probability of incorrect conclusions are most commonly applied in confirmatory trials. However, as stressed above, it is important to address multiplicity effects in exploratory trials (phase 1 and 2 trials). Methods for an indirect treatment of multiplicity have been successfully applied in exploratory trials (e.g., the development of dose-finding algorithms in phase 2 trials).[13]

Multiplicity also arises in the analysis of safety measures, including adverse events. Safety analyses are often viewed as exploratory, and P values used in summaries of adverse events are typically treated as "flagging devices." Statistical methods, such as the double false discovery rate method, can be applied to support a more rigorous assessment of adverse events that accounts for inherent multiplicity.[14]

MULTIPLICITY ADJUSTMENTS

Numerous statistical methods for performing adjustments for multiplicity, known as multiple testing procedures or multiple tests, have been developed.[15] These methods have been broadly applied in phase 3 trials to control the overall type I error rate at the prespecified 5% level.

CLINICAL TRIALS WITH A SINGLE SOURCE OF MULTIPLICITY

The choice of multiple tests in a particular trial is driven by the available clinical or historical information and the available statistical information. Clinical information is often used to define relevant dependencies among the individual objectives (e.g., to specify a hierarchical sequence for the end points from the most important to the least important in a phase 3 trial). If a meaningful ordering of the end points cannot be prespecified, they will be examined in a data-driven sequence (e.g., the end points of interest will be tested beginning with the least significant one or the most significant one). Multiple tests used in these two settings are said to rely on a predefined testing sequence or a data-driven testing sequence.

Information on statistical features of a study design or analytic strategy (e.g., information on the joint distribution of the end points or dose–control comparisons) drives the selection of efficient multiple tests. Basic tests known as nonparametric tests, such as the Bonferroni test, can be applied to any multiplicity problem but are generally inefficient — in the sense that they

Table 1. Statistical Methods Commonly Used in Clinical Trials to Correct for Multiplicity.

Test Classification	Examples	Key Properties and Applications*
Tests with a predefined testing sequence based on clinical information	Fixed-sequence test	A fixed-sequence test can be applied to any multiplicity problem when the clinical objectives (e.g., end points) can be arranged in a meaningful way. This test tends to produce spurious results if the prespecified ordering is based on unreliable information.
Tests with a data-driven testing sequence based on clinical information		
Nonparametric tests	Bonferroni, Holm, and fallback tests	Nonparametric tests can be applied to any multiplicity problem but tend to be inefficient and lead to a lower overall probability of success.
Semiparametric tests	Hochberg and Hommel tests	Semiparametric tests are more efficient than nonparametric tests and are commonly used in clinical trials with several dose–control comparisons and patient populations.
Parametric tests	Regular Dunnett and stepdown Dunnett tests	Parametric tests are more efficient than nonparametric and semiparametric tests but rely on very specific statistical assumptions (e.g., correlations must be known). Parametric tests are used in clinical trials with several dose–control comparisons and patient populations.

* Efficiency in terms of nonparametric, semiparametric, and parametric tests refers to an increased or a reduced probability of a significant treatment effect in a trial. Inefficient multiple tests result in a lower probability of establishing a significant treatment effect than efficient multiple tests.

overcorrect for multiplicity and lead to a lower probability of establishing a significant treatment effect — as compared with semiparametric or parametric tests. Semiparametric or parametric tests are more efficient and can be applied when additional statistical information is available (e.g., the dose–control comparisons are positively correlated, or a meaningful model can be formulated for the outcome).

A description of these classes of tests is provided in Table 1. Tests commonly used to adjust for multiple end points and for multiple patient populations are illustrated below. For further information, see the Supplementary Appendix, available with the full text of this article at NEJM.org.

CLINICAL TRIALS WITH SEVERAL SOURCES OF MULTIPLICITY

Several primary and secondary end points are commonly used to provide a comprehensive characterization of the efficacy properties of new treatments.[16,17] Clinical end points are often grouped into families of primary, secondary, and exploratory end points.[7,9] Demonstration of a significant treatment effect on the primary end points is essential for gaining regulatory approval, and the secondary end points may support additional regulatory claims. These end points are presented in the information on the product label, supported by inferential statements (i.e., they can be accompanied by P values or confidence intervals). Exploratory end points may provide general supportive evidence of effectiveness, but unlike primary and secondary end points, they can be included in the information on the product label only for purely descriptive purposes (i.e., without the support of statistical inferences).

Tests that are specifically designed for handling complex multiplicity problems with several primary and secondary end points are termed gatekeeping tests.[5,18] By controlling the number of comparisons that are considered valid, gatekeeping tests enable clinical trial sponsors to provide additional efficacy information on the product label that will be useful for prescribers and patients. Gatekeeping tests are discussed below in the section on clinical trials with several sources of multiplicity.

MULTIPLE END POINTS

Complex causes of autoimmune, cardiovascular, and other diseases often necessitate multiple primary end points to accurately describe a therapeutic benefit.[19] In addition, it is customary to predefine a set of secondary end points in phase 3 trials in order to characterize the efficacy of new treatments, create differentiating factors, and strengthen the information on the product label. In both cases, multiplicity is induced by multiple opportunities to claim success in a trial, as shown in the examples below.

CLINICAL TRIALS WITH A SINGLE SOURCE OF MULTIPLICITY

Multiplicity adjustment strategies in trials with a single source of multiplicity based on primary end points can be illustrated with a phase 3 trial for the treatment of metastatic castration-resistant prostate cancer.[20] This trial was conducted to evaluate survival benefits associated with abiraterone plus prednisone versus prednisone alone. The trial's primary objective was formulated in terms of radiographic progression-free survival and overall survival. Because there were two potential efficacy claims, the Bonferroni test was used to preserve the overall type I error rate at a two-sided alpha level of 0.05. The overall error rate was split unequally between the end points. Since progression-free survival data mature faster than overall survival data, the improvement in radiographic progression-free survival would be declared significant at P≤0.01. This analysis was to be followed by the overall survival analysis, which would be significant at P≤0.04. Similar alpha-splitting strategies have been used in other oncology trials, including the PREVAIL trial,[21] which also evaluated treatment for prostate cancer.

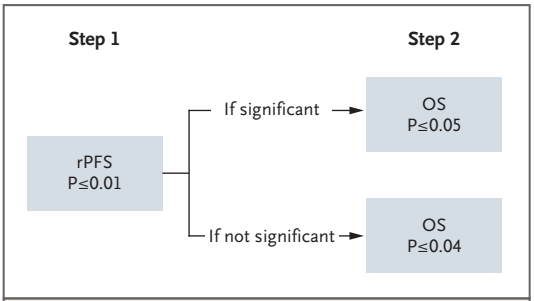

Figure 1. Decision Rules of the Fallback Test with Unequal Alpha Allocation in a Phase 3 Oncology Trial with Two Primary End Points.

The two primary end points are radiographic progression-free survival (rPFS) and overall survival (OS). The arrows represent the decision paths supported by the fallback test. In step 1, a significant treatment effect on rPFS is established at P≤0.01. In step 2, a significant treatment effect on OS is established if rPFS was significant in step 1 and P≤0.05 for OS or if rPFS was not significant in step 1 but P≤0.04 for OS.

Although the basic Bonferroni test certainly controls the error rate in clinical trials with several end points, it is the most conservative multiplicity adjustment and it can be improved in several ways. The fallback test serves as a simple extension of the Bonferroni test. As shown in Figure 1, this test has two steps. In step 1, the fallback test uses the same decision rule for radiographic progression-free survival as the Bonferroni test (e.g., the between-group difference in radiographic progression-free survival is significant at P≤0.01). However, the chances of establishing a survival benefit can be improved in step 2 if the effect on progression-free survival is significant. The analysis of overall survival is performed at a higher level if the treatment effect on radiographic progression-free survival is significant in step 1 (i.e., P≤0.05). Otherwise, the analysis in step 2 is performed at the same level as in the Bonferroni test (i.e., P≤0.04).

An important feature of oncology trials is that an analysis of progression-free survival is performed before the analysis of overall survival. If the primary end points are evaluated at the same time, the fallback test displayed in Figure 1 can be replaced by the Holm test, which is more efficient. The testing strategy, shown in Figure 2, is set up for a trial with two primary end points (end point 1 and end point 2). Both the fallback and Holm tests are easily extended to trials with three or more end points.

When alpha-splitting methods are applied in clinical trials, the actual alpha allocation is determined in an ad hoc fashion. The allocation can be selected to maximize the overall probability of success in the trial with the use of standard clinical-trial optimization methods.[22]

CLINICAL TRIALS WITH SEVERAL SOURCES OF MULTIPLICITY

Advanced gatekeeping tests are used in clinical trials with two or three sources of multiplicity due to several primary and secondary end points,

several dose–placebo comparisons, or other factors. The lurasidone program for the treatment of schizophrenia[23,24] provides examples of challenging multiplicity problems of this kind. This program included trials for evaluating the efficacy profile of two or three doses of lurasidone versus placebo on the basis of a single primary end point (change from baseline in the total score on the Positive and Negative Syndrome Scale at week 6) and two secondary end points (change from baseline in the Clinical Global Impression of Severity score at week 6 and change from baseline in the total score on the Positive and Negative Syndrome Scale at day 4). The resulting assessments were grouped into families as shown in Figure 3. It is important to point out that the gatekeeping tests accounted for the clinically relevant dependencies among the individual objectives.

An efficient gatekeeping test that applied the Hommel-type tests within each family was developed to control the overall type I error rate. Defining the decision rules used in the Hommel-based gatekeeping test is beyond the scope of this review, but the test is easy to implement in phase 3 trials.[25] General principles for constructing powerful gatekeeping tests in clinical trials with secondary end points can be applied to a broad class of advanced multiplicity problems.[26,27]

COMPOSITE END POINTS

Multiplicity issues also arise in clinical trials that use composite end points (e.g., major adverse cardiac events in trials of treatment for cardiovascular disease).[19] When analytic strategies for composite end points are defined, it is important to describe how each component's effect on the overall conclusions will be interpreted. If a disproportional influence of a particular component of an end point has been established, it may lead to a decision to restrict the product label in order to focus on this specific component rather than on the composite end point. Such a decision was made, for example, in the LIFE (Losartan Intervention for Endpoint Reduction in Hypertension) trial because losartan had a major effect on the blood pressure and stroke components of the composite end point but a smaller effect on the myocardial infarction component.[28] Several rules for evaluating the effect of individual components, including "soft" components such as recurrent angina and "hard" components such as mortality, in a thrombolytic clinical trial, have been proposed in the literature.[29]

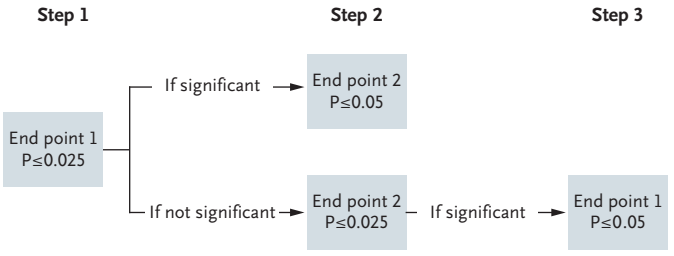

Figure 2. Decision Rules of the Holm Test with Equal Alpha Allocation in a Phase 3 Trial with Two Primary End Points.

The arrows represent the decision paths supported by the Holm test. In step 1, a significant treatment effect on end point 1 is established at P≤0.025. In step 2, a significant treatment effect on end point 2 is established if the treatment effect on end point 1 was significant in step 1 and P≤0.05 for end point 2 or if end point 1 was not significant in step 1 but P≤0.025 for end point 2. Finally, in step 3, end point 1 can be examined again if it was not significant in step 1. A significant treatment effect for end point 1 is established if the treatment effect for end point 2 was significant in step 2 and P≤0.05 for end point 1.

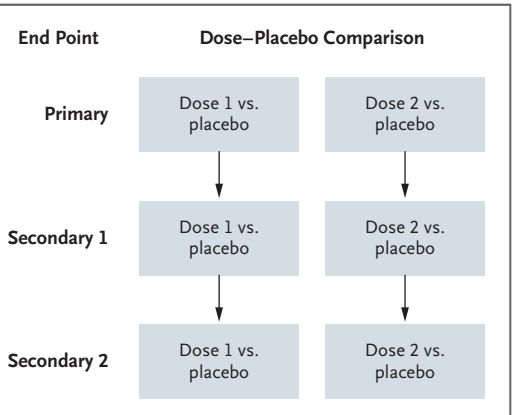

Figure 3. Testing Strategy Used in a Phase 3 Trial of Lurasidone versus Placebo, with Three End Points and Two Doses.

The arrows represent the decision paths, which were consistent with the clinically relevant dependencies. A dose–placebo comparison for a secondary end point could be performed only if a significant effect on the primary end point had been established at the particular dose.

MULTIPLE PATIENT POPULATIONS

Approaches to performing subgroup analysis in clinical trials can be viewed as either confirma-

Table 2. Use of the Fixed-Sequence and Hochberg Tests to Evaluate the Treatment Effect in the Three Predefined Populations in the APEX Trial.*

Population	P Value (Betrixaban vs. Enoxaparin)	Fixed-Sequence Test	Hochberg Test
Subpopulation 1	0.054	No significant effect	No significant effect
Subpopulation 2	0.03	No significant effect	No significant effect
All-comers population	0.006	No significant effect	Significant effect

* APEX denotes Acute Medically Ill VTE (Venous Thromboembolism) Prevention with Extended Duration Betrixaban.

tory or exploratory.[30] Confirmatory subgroup analysis is applicable to late-stage clinical trials and involves evaluation of the therapeutic effect in several prospectively defined subsets of the trial population.[31] Exploratory subgroup analysis is aimed at discovering new features of a treatment's efficacy or safety profile in an ad hoc manner. This section focuses on confirmatory analysis.

As interest in the development of targeted therapies has increased, numerous phase 3 trials have had multiple-population designs. In these trials, the efficacy of a new treatment has been studied in two or more prespecified subpopulations in addition to the intention-to-treat population, also known as the all-comers population. An example is SATURN (Sequential Tarceva in Unresectable NSCLC), which evaluated the role of erlotinib maintenance therapy in patients who had lung cancer with or without a mutation in the epidermal growth factor receptor, the target of erlotinib.[32] Multiplicity occurred in this trial because of the possibility of making an efficacy claim in the intention-to-treat population and a predefined subpopulation.

When a testing strategy is developed for a trial with several populations, it is critical to ensure that the strategy is flexible and treats the predefined populations as interchangeable rather than hierarchically ordered. The APEX (Acute Medically Ill VTE [Venous Thromboembolism] Prevention with Extended Duration Betrixaban) trial[33] can be used to compare different approaches to multiplicity adjustments in multiple-population trials. This trial investigated the advantages of betrixaban as compared with enoxaparin in patients at risk for venous thrombosis. The primary analysis was performed in the all-comers population and two target subpopulations (subpopulations 1 and 2). The two-sided P values for the between-group differences in the three populations are shown in Table 2.

Figure 4 shows the decision rules for the fixed-sequence test, which relies on a rigid testing strategy with hierarchically ordered populations, and Figure 5 shows the decision rules for the Hochberg test, which uses a flexible testing sequence. Table 2 applies the decision rules and provides a comparison of the two tests in the APEX trial. When the fixed-sequence test is carried out, the first P value in the sequence is greater than 0.05 (P=0.054). Owing to this result, no efficacy claim can be formulated for betrixaban in any of the patient populations. The analyses in subpopulation 2 and the all-comers population can only be considered exploratory, despite the fact that the between-group differences in these two populations are significant at a 5% level (P=0.03 and P=0.006). This particular example illustrates an important property of the fixed-sequence test: it can be justified only if the most significant P value is expected in sub-

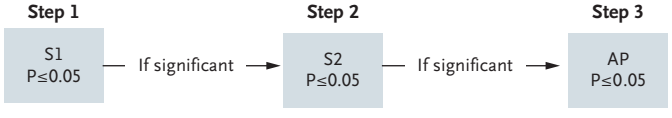

Figure 4. Decision Rules of the Fixed-Sequence Test in the APEX Trial.
This test relies on a predefined sequence: subpopulation 1 (S1) is to be tested first, followed by subpopulation 2 (S2), followed by the all-comers population (AP). In step 1, a significant treatment effect was established in S1 at P≤0.05. In step 2, a significant treatment effect was established in S2 if the effect in S1 was significant in step 1 and P≤0.05. In Step 3, a significant treatment effect was established in AP if the effect in S2 was significant in step 2 and P≤0.05. APEX denotes Acute Medically Ill VTE [Venous Thromboembolism] Prevention with Extended Duration Betrixaban.

population 1 and the least significant P value is expected in the all-comers population. If this assumption is not met, as was the case in the APEX trial, this multiple test is likely to miss statistically and clinically relevant results.

As shown in Figure 5, a data-driven testing sequence is used when the Hochberg test is carried out. Table 2 shows that, with this multiple test, betrixaban would have been declared superior to enoxaparin in the all-comers population, since P≤0.017. Multiple tests with a data-driven sequence support more flexible decision paths, as compared with the fixed-sequence test. For example, these tests enable the trial's sponsor to claim efficacy in one patient population (e.g., the all-comers population) even if the treatment effect in another population (e.g., subpopulation 1) is not significant.

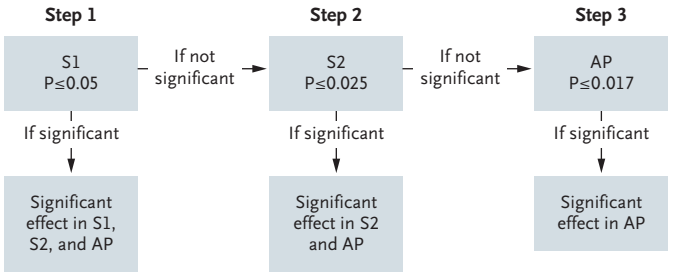

Figure 5. Decision Rules of the Hochberg Test in the APEX Trial.

This test relies on a data-driven testing sequence (i.e., S1 is tested first because it corresponds to the largest P value, followed by S2 with the second largest P value, followed by AP with the smallest P value). In step 1, a significant treatment effect was established in S1, S2, and AP at P≤0.05. In step 2, a significant treatment effect was established in S2 and AP if the effect in S1 was not significant in step 1 but P≤0.025. In step 3, a significant treatment effect was established in AP if the effect in S2 was not significant in step 2 but P≤0.017.

CONCLUSIONS

Virtually all confirmatory phase 3 trials are designed to pursue multiple clinical objectives that are formulated on the basis of several end points or doses of an experimental treatment. Interpretation of clinical trial results may be quite complicated in the presence of multiplicity, since it often increases the chances of drawing an incorrect conclusion. A large number of statistical strategies, known as multiplicity adjustments, have been developed to address different sources of multiplicity and to control the probability of erroneously concluding that the experimental treatment is effective.

The choice of a multiplicity adjustment is likely to have a major effect on the overall conclusions in a trial. It is critical to consider relevant clinical and statistical information and perform a comprehensive review of all applicable multiplicity adjustment strategies in order to identify a strategy that is aligned with the trial's objectives and maximizes the probability of success. This includes information on the relative importance of the end points or patient populations and key statistical features such as correlations among the test statistics in a multiplicity problem. Extensive clinical trial simulations are often conducted to facilitate the process of selecting the most efficient and robust multiplicity adjustment for a particular clinical trial.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

REFERENCES

1. Bauer P. Multiple testing in clinical trials. Stat Med 1991;10:871-89.
2. D'Agostino RB, Heeren TC. Multiple comparisons in over-the-counter drug clinical trials with both positive and placebo controls. Stat Med 1991;10:1-6.
3. D'Agostino RB, Massaro J, Kwan H, Cabral H. Strategies for dealing with multiple treatment comparisons in confirmatory clinical trials. Drug Inf J 1993;27:625-41.
4. Sankoh AJ, Huque MF, Dubey SD. Some comments on frequently used multiple endpoint adjustment methods in clinical trials. Stat Med 1997;16:2529-42.
5. Dmitrienko A, D'Agostino RB Sr, Huque MF. Key multiplicity issues in clinical drug development. Stat Med 2013;32:1079-111.
6. Huque MF, Dmitrienko A, D'Agostino RB. Multiplicity issues in clinical trials with multiple objectives. Stat Biopharm Res 2013;5:321-37.
7. Food and Drug Administration. Multiple endpoints in clinical trials: guidance for industry. January 2017 (https://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm536750.pdf).
8. European Medicines Agency. Points to consider on multiplicity issues in clinical trials. September 2002 (http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003640.pdf).
9. European Medicines Agency. Guideline on multiplicity issues in clinical trials. December 2016 (http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2017/03/WC500224998.pdf).
10. Dmitrienko A, D'Agostino R Sr. Traditional multiplicity adjustment methods in clinical trials. Stat Med 2013;32:5172-218.
11. Offen W, Chuang-Stein C, Dmitrienko A, et al. Multiple co-primary endpoints: medical and statistical solutions: a report from the multiple endpoints expert team of the Pharmaceutical Research and Man-

ufacturers of America. Drug Inf J 2007;41:31-46.
12. Chuang-Stein C, Dmitrienko A, Offen W. Discussion of "Some controversial multiple testing problems in regulatory applications." J Biopharm Stat 2009;19:14-21.
13. Bretz F, Pinheiro JC, Branson M. Combining multiple comparisons and modeling techniques in dose-response studies. Biometrics 2005;61:738-48.
14. Mehrotra DV, Heyse JF. Use of the false discovery rate for evaluating clinical safety data. Stat Methods Med Res 2004;13:227-38.
15. Dmitrienko A, Bretz F, Westfall PH, et al. Multiple testing methodology. In: Dmitrienko A, Tamhane AC, Bretz F, eds. Multiple testing problems in pharmaceutical statistics. New York: Chapman and Hall/CRC Press, 2009:35-98.
16. D'Agostino RB Sr. Controlling alpha in a clinical trial: the case for secondary endpoints. Stat Med 2000;19:763-6.
17. Huque MF, Röhmel J. Multiplicity problems in clinical trials: a regulatory perspective. In: Dmitrienko A, Tamhane AC, Bretz F, eds. Multiple testing problems in pharmaceutical statistics. New York: Chapman & Hall/CRC Press, 2010:1-34.
18. Dmitrienko A, Tamhane AC. Gatekeeping procedures in clinical trials. In: Dmitrienko A, Tamhane AC, Bretz F, eds. Multiple testing problems in pharmaceutical statistics. New York: Chapman and Hall/CRC Press, 2009:165-91.
19. Sankoh AJ, Li H, D'Agostino RB Sr. Use of composite endpoints in clinical trials. Stat Med 2014;33:4709-14.
20. Ryan CJ, Smith MR, de Bono JS, et al. Abiraterone in metastatic prostate cancer without previous chemotherapy. N Engl J Med 2013;368:138-48.
21. Beer TM, Armstrong AJ, Rathkopf DE, et al. Enzalutamide in metastatic prostate cancer before chemotherapy. N Engl J Med 2014;371:424-33.
22. Dmitrienko A, Pulkstenis E, eds. Clinical trial optimization using R. New York: Chapman and Hall/CRC Press, 2017.
23. Meltzer HY, Cucchiaro J, Silva R, et al. Lurasidone in the treatment of schizophrenia: a randomized, double-blind, placebo- and olanzapine-controlled study. Am J Psychiatry 2011;168:957-67.
24. Nasrallah HA, Silva R, Phillips D, et al. Lurasidone for the treatment of acutely psychotic patients with schizophrenia: a 6-week, randomized, placebo-controlled study. J Psychiatr Res 2013;47:670-7.
25. Brechenmacher T, Xu J, Dmitrienko A, Tamhane AC. A mixture gatekeeping procedure based on the Hommel test for clinical trial applications. J Biopharm Stat 2011;21:748-67.
26. Dmitrienko A, Tamhane AC. Mixtures of multiple testing procedures for gatekeeping applications in clinical trials. Stat Med 2011;30:1473-88.
27. Dmitrienko A, Tamhane AC. General theory of mixture procedures for gatekeeping. Biom J 2013;55:402-19.
28. Dahlöf B, Devereux RB, Kjeldsen SE, et al. Cardiovascular morbidity and mortality in the Losartan Intervention For Endpoint reduction in hypertension study (LIFE): a randomised trial against atenolol. Lancet 2002;359:995-1003.
29. Huque MF, Alosh M, Bhore R. Addressing multiplicity issues of a composite endpoint and its components in clinical trials. J Biopharm Stat 2011;21:610-34.
30. Lipkovich I, Dmitrienko A, D'Agostino RB Sr. Tutorial in biostatistics: data-driven subgroup identification and analysis in clinical trials. Stat Med 2017;36:136-96.
31. Millen BA, Dmitrienko A, Ruberg S, Shen L. A statistical framework for decision making in confirmatory multipopulation tailoring clinical trials. Drug Inf J 2012;46:647-56.
32. Cappuzzo F, Ciuleanu T, Stelmakh L, et al. SATURN: a double-blind, randomized, phase III study of maintenance erlotinib versus placebo following nonprogression with first-line platinum-based chemotherapy in patients with advanced NSCLC. J Clin Oncol 2009;27:8001. abstract.
33. Cohen AT, Harrington RA, Goldhaber SZ, et al. Extended thromboprophylaxis with betrixaban in acutely ill medical patients. N Engl J Med 2016;375:534-44.

IMAGES IN CLINICAL MEDICINE

The *Journal* welcomes consideration of new submissions for Images in Clinical Medicine. Instructions for authors and procedures for submissions can be found on the *Journal*'s website at NEJM.org. At the discretion of the editor, images that are accepted for publication may appear in the print version of the *Journal*, the electronic version, or both.