SUPPLEMENT ARTICLE

Forum Report: Issues in the Design of Trials of Drugs for the Treatment of Invasive Aspergillosis

John E. Bennett,[1] John Powers,[3] Ben de Pauw,[8] William Dismukes,[4] John Galgiani,[5] Michel Glauser,[9] Raoul Herbrecht,[10] Carol Kauffman,[6] Jeannette Lee,[4] Peter Pappas,[4] John Rex,[7] Paul Verweij,[8] Claudio Viscoli,[11] and Thomas Walsh[2,a]

[1]Clinical Mycology Section, Laboratory of Clinical Investigation, National Institute of Allergy and Infectious Diseases, and [2]National Cancer Institute, National Institutes of Health, Bethesda, and [3]Division of Special Pathogen and Immunologic Drug Products, US Food and Drug Administration, Rockville, Maryland; [4]University of Alabama at Birmingham; [5]VA Medical Center and University of Arizona, Tucson; [6]VA Medical Center and University of Michigan, Ann Arbor; [7]University of Texas Medical School, Houston; [8]University Hospital St. Radboud, Nijmegen, Netherlands; [9]University Hospital, Lausanne, Switzerland; [10]Hôpital de Hautepierre, Strasbourg, France; and [11]University of Genoa, National Institute for Cancer Research, Italy.

A recent trial of drugs for invasive aspergillosis was used as a background for discussing critical features in the design of antifungal trials. The study under discussion allowed stopping either drug without classifying the patient as having treatment failure, so the trial should be understood as a comparison of 2 treatment strategies, not just 2 drugs. Although the study was a noninferiority trial, the outcome permitted a claim of superiority. Use of the category of "probable" in addition to "proven" aspergillosis permitted inclusion of patients for whom the diagnosis was less certain but who were still early enough in the disease progression to respond to therapy. Different opinions still exist about some of the criteria for the diagnosis of "probable" aspergillosis. A blinded data review committee was helpful in evaluating efficacy in this unblinded trial but had limited value in assessing toxicity. An understanding of these features of design of antifungal drug trials is important in applying the results to clinical practice.

Previous publications about drug efficacy in invasive aspergillosis have been clinical descriptions or, in the one prospective clinical trial, a comparison of 2 doses of the same drug [1]. Definitions of invasive aspergillosis in prior drug studies have been varied and often imprecise. The recent multicenter comparative study of voriconazole versus amphotericin B for the treatment of invasive aspergillosis [2] used detailed descriptions for case definitions and presented an opportunity to discuss the study entry criteria, end points, the selection of comparators, the use of data review committees (DRCs), statistical analysis, and population pharmacokinetics.

STUDIES THAT PERMIT A CHANGE FROM THE RANDOMIZED DRUG

The voriconazole study permitted a change from the randomized drug to any other licensed antifungal drug, without requiring that the patient be classified as having had treatment failure of the randomized drug. This strategy is similar to the "success with modification" used previously in studies by the National Cancer Institute of empirical antibacterial therapy in neutropenic patients. The Immunocompromised Host Society rejected the success-with-modification design after a consensus meeting. Viscoli [3] distinguished between clinical trials that were "pragmatic" and those that were

[a] Additional participants are listed after the text.

Reprints or correspondence: Dr. John E. Bennett, Clinical Mycology Section, Laboratory of Clinical Investigation, Infectious Diseases, Building 10, Room 11C304, National Institutes of Health, Bethesda, MD 20892 (jbennett@niaid.nih.gov).

Clinical Infectious Diseases 2003; 36(Suppl 3):S113–6

1058-4838/2003/3608S3-0003$15.00

"explanatory." According to Viscoli's definition, pragmatic trials such as the voriconazole trial or the success-with-modification trials are suitable to evaluate a management strategy and can apply to clinical practice but cannot give a quantitative estimate of the efficacy of a new drug compared with that of an old drug. Measurement of mortality and detailed definitions of permissible therapeutic modifications are essential components of pragmatic clinical trials. The case report form should provide a complete explanation of the reasons for any change in therapy. As discussed below, a blinded DRC can validate whether the therapeutic modifications were permissible within the protocol.

SELECTION OF COMPARATOR DRUGS

The investigators and the sponsor of the trial of voriconazole in treating invasive aspergillosis selected amphotericin B deoxycholate as the initial comparator drug because it was the only licensed drug for the initial treatment of invasive aspergillosis in the United States. The toxicity of amphotericin B deoxycholate led to a change to another licensed antifungal drug for 107 of 133 patients in the comparator arm of the trial. Although amphotericin B deoxycholate has been accepted for many years as the standard for primary treatment of invasive aspergillosis, the advent of new lipid-based formulations of amphotericin B and the echinocandins raised the question of whether it is still appropriate to use amphotericin B deoxycholate as the comparator in all new antifungal trials. The answer hinges on whether a different formulation of an approved drug, such as a lipid amphotericin B formulation, can be accepted as a comparator when it is not approved by the US Food and Drug Administration (FDA) as initial therapy for invasive aspergillosis. Accepting an unlicensed formulation as the comparator would be tantamount to approving the formulation without a clinical trial, creating an unacceptable precedent. At the meeting of the FDA Advisory Committee on Caspofungin, the FDA indicated willingness to consider alternative comparators at the protocol design stage, as long as investigators and sponsors provided further supporting data on the efficacy of comparators unlicensed for that indication. The FDA also stated that it would consider the results of a clinical trial designed to show the superiority (not equivalence) of a new drug compared with an unlicensed drug or formulation. This is based on the fact that a superiority trial that shows that a particular drug is indeed superior to another proves that the test drug must be better than placebo. This is a high hurdle for a new drug, however, and investigators and sponsors rarely choose this option. There is no FDA requirement that a new drug be superior to already licensed drugs in either efficacy or toxicity, although there are marketing and clinical reasons for wanting such an advantage.

NONINFERIORITY TRIALS

The voriconazole aspergillosis study was a noninferiority trial, like most trials of new antifungal drugs, with a noninferiority margin (delta) of 20%. That is, the study hoped to reject the hypothesis that the voriconazole arm experienced a 20% worse outcome than the amphotericin B deoxycholate arm. It is important to understand the difference between this kind of a design and a superiority trial. In a superiority trial, the study is designed to determine whether a particular drug is superior to another. Failure to establish superiority of one drug over the other does not definitely establish that they are equivalent, because the sample size of the trial may be too small to establish noninferiority. A potentially useful equivalent drug could be discarded as a result of a failed superiority trial. In a noninferiority trial, one drug is assumed to be inferior to another drug and the study is designed to establish whether 2 drugs are equally effective. The typical criterion for noninferiority between drugs is to establish limits of a confidence interval for the difference in efficacy between drugs. If the confidence interval lies within prespecified limits, the drugs are considered comparable with respect to efficacy. Without prior knowledge that the older drug is effective, one cannot conclude from a noninferiority trial that either drug is better than nothing at all. Although a noninferiority trial is not designed to test superiority, that conclusion can be inferred if the lower boundary of the confidence interval around the difference between the response rates of the test drug minus the comparator drug is >0 [4].

Once the older, comparator drug has been shown to be superior to no treatment, the next consideration in selecting a noninferiority margin is the clinical judgment of what is an acceptable loss of efficacy relative to current therapies for that particular disease. This must also take into account that selecting a smaller noninferiority margin may increase the sample size for a clinical trial. For severe diseases, selection of a relatively small delta is preferable, to ensure as little loss of efficacy as possible relative to the control agent. This may not be possible if the disease is relatively uncommon. In some cases, the number of patients required to allow the trial to have adequate statistical power may be larger than the number of patients who have the illness in question. However, the noninferiority margin cannot be greater than the benefit of drug therapy over placebo. For instance, if the benefit of drug therapy over placebo is 15%, then a noninferiority margin of 10% would be appropriate whereas a noninferiority margin of 20% would not.

In the trial of voriconazole to treat invasive aspergillosis, the noninferiority margin selected before initiation of the trial was 20% [2]. This means that investigators could conclude that the voriconazole arm was not inferior to the amphotericin B arm if the lower bound of the 95% confidence interval around the

difference in the efficacy of voriconazole minus the difference in the control regimen was not more negative than −20%. The noninferiority margin of 20% was justified by historical data indicating that the efficacy of amphotericin B in this population was well above 20%. Because the lower bound of the 95% confidence interval around the difference between the 2 drugs (voriconazole minus amphotericin B followed by other licensed therapy) for the primary end point was >0, it was possible to conclude that the voriconazole arm was statistically superior to the amphotericin B deoxycholate arm.

ENTRY CRITERIA

Efforts to reach a consensus about entry into a therapeutic trial for deep mycoses, including aspergillosis, in immunocompromised patients have resulted in a publication [5]. To date, these criteria have not been used in a clinical trial. Although there are some differences between the entry criteria in the voriconazole trial and the published consensus, the 2 sets of diagnostic criteria do agree in including the important element of combining the host's risk factors with the radiological and laboratory findings.

USE OF THE HALO SIGN IN DIAGNOSIS OF INVASIVE ASPERGILLOSIS

In prolonged neutropenia, investigators have advocated the identification of a halo sign on thoracic CT as a sensitive, early sign of invasive pulmonary aspergillosis [6]. The specificity of the halo sign is less clear. There is no consensus about the precise radiological definition of the halo sign so that one would expect variability between centers in detecting this finding. In the voriconazole randomized trial of invasive aspergillosis, the DRC radiologists could not confirm the presence of a halo sign for 57 of 95 patients. The major problem in confirming a diagnosis appeared to be the lack of a uniform definition in hospital radiology departments. Lack of appropriate documentation may also have been a significant contributing factor. It is possible that the DRC did not receive the best CT scans to demonstrate the halo. Because a uniform definition was used among the radiologists in the voriconazole DRC, there was complete agreement in the radiological assessment of response between the United States and European DRCs in >80% of cases. The therapeutic outcomes of patients for whom the diagnosis of invasive aspergillosis was rejected by the DRC radiologists were the same as outcomes for those whose diagnosis was accepted by the DRC, suggesting, but not confirming, a similar etiology in both groups. If this is true, then the sensitivity of the halo sign may be less than the reported figure of 84% (16/19 proven cases) [6].

A uniform definition of the halo sign is needed and should include radiological criteria, a description of the patient population to which the sign applies, estimates of the sensitivity and specificity of the sign, and a description of other diseases in this same population that can produce a halo sign.

DRCS

Blinded DRCs examine radiographs and case report forms that have been purged of details that may disclose the identity of the randomized drug. The trial of voriconazole to treat aspergillosis [2] included both US and European DRCs composed of clinicians and radiologists who confirmed study eligibility and response to therapy. As exemplified in the voriconazole trial, blinded DRCs can play a valuable role in the successful conduct of randomized multicenter trials by identifying a need for additional data about individual patients, confirming eligibility for study entry, validating outcome, and providing a central place for resolving technical issues as they arise. In addition, patients with unusual outcomes, such as unexpected treatment failure, are identified and given special scrutiny. Blinded DRCs also ensure that rules are applied uniformly in both arms of the trial. This is particularly important in complex trials that include some inexperienced investigators.

Blinded DRCs are, however, not without some disadvantages. Because of data purging, DRC members frequently have an incomplete data set to review. The DRC cannot adequately assess drug toxicity because laboratory values and narrative comments that might provide a clue to the drug identity have been removed. Although the DRC can review toxicity once the study is completed and the blind has been broken, toxicity data not already reported to the company are difficult to retrieve after such a long time lag. A blinded DRC may reverse the investigator's claim of efficacy simply because there is insufficient documentation provided to support it. When the DRC deletes cases from the modified intent-to-treat analysis, the reduced patient numbers decrease the power of the study and may limit the opportunity to draw valid conclusions. The FDA routinely examines analyses of both the intent-to-treat population and the modified intent-to-treat population to see whether the results are in accordance. Studies that have a large number of patients excluded from the modified intent-to-treat analysis might show dissimilar results from the intent-to-treat population because of some bias in the DRC exclusion process. When >1 DRC is used, evidence must be provided that the results are in agreement. In the trial of voriconazole to treat aspergillosis, comparison of the results of the European and US DRC evaluations of the same case report forms and radiological studies found >90% agreement as to outcome as suc-

cess or failure, indicating that both DRCs were applying the same criteria.

ADDITION OF NARRATIVES TO CASE REPORT FORMS

Improvements in the design of case report forms would enhance the work of blinded DRCs. Specifically, the addition of narrative patient summaries, written by the investigator, would greatly improve the DRC's understanding of why investigators made certain clinical decisions. Experience from the DRCs of the trial of voriconazole as empirical antifungal therapy and the trial of caspofungin as treatment for invasive aspergillosis showed that asking the investigator to prepare individual patient summaries can help a blinded DRC to understand more about the time course and rationale for clinical decisions. The summaries may also provide valuable information about why a drug was stopped for lack of efficacy. Improved, more complete check-off boxes on case report forms may also help the blinded DRC to confirm or disagree that early discontinuation of therapy for lack of efficacy or toxicity was appropriate.

The caveats to the use of narratives in the case report form include the necessity for keeping the DRC blinded as to study drug and the possibility that the study nurses who fill out the majority of the case report forms may have difficulty writing a summary that captures all of the major management decisions. The addition of narratives does add considerably to the workload of the data managers and statisticians. Use of narrative summaries also raises questions on how best to analyze these additional data.

Additional Participants

Alice Baruch (Pfizer Pharmaceuticals Group, New York), Helen Boucher (Pfizer Global Research and Development, New London, CT), Donald Buell (Fujisawa Healthcare, Northbrook, IL), John Edwards (Harbor/UCLA Medical Center, Torrance, CA), Christopher Hitchcock (Pfizer Global Research and Development, Sandwich, UK), Michael Hodges (Pfizer Global Research and Development, New London, CT), Ann Kolokathis (Pfizer, New York), Helene Panzer (Pfizer Pharmaceuticals Group, New York), Thomas Patterson (University of Texas Health Science Center, San Antonio), John Perfect (Duke University Medical Center, Durham, NC), Patricia Ribaud (Hôpital Saint-Louis, Paris), Jack Sobel (Wayne State University School of Medicine, Detroit), Tania Sorrell (Westmead Hospital, Westmead, Australia).

References

1. Ellis M, Spence D, de Pauw B, et al. An EORTC international multicenter randomized trial (EORTC number 19923) comparing two dosages of liposomal amphotericin B for treatment of invasive aspergillosis. Clin Infect Dis **1998**; 27:1406–12.
2. Herbrecht R, Denning D, Patterson T, et al. Voriconazole versus amphotericin B for primary therapy of invasive aspergillosis. N Engl J Med **2002**; 347:408–15.
3. Viscoli C. The evolution of the empirical management of fever and neutropenia in cancer patients. J Antimicrob Chemother **1998**; 41(suppl D):65–80.
4. Powers JH, Ross DB, Brittain E, et al. The United States Food and Drug Administration and noninferiority margins in clinical trials of antimicrobial agents. Clin Infect Dis **2002**; 34:879–81.
5. Ascioglu S, Rex JH, de Pauw B, et al. Defining opportunistic invasive fungal infections in immunocompromised patients with cancer and hematopoietic stem cell transplants: an international consensus. Clin Infect Dis **2002**; 34:7–14.
6. Caillot D, Casasnovas O, Bernard A, et al. Improved management of invasive pulmonary aspergillosis in neutropenic patients using early thoracic computed tomographic scan and surgery. J Clin Oncol **1997**; 15: 139–47.