Forum Report: Issues in the Evaluation of Diagnostic Tests, Use of Historical Controls, and Merits of the Current Multicenter Collaborative Groups

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

[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]VA Medical Center and University of Michigan, Ann Arbor; [5]University of Alabama at Birmingham; [6]VA Medical Center and University of Arizona, Tucson; [7]University of Texas Medical School, Houston; [8]University Hospital St. Radboud, Nijmegen, The Netherlands; [9]University Hospital, Lausanne, Switzerland; [10]Hôpital de Hautepierre, Strasbourg, France; and [11]University of Genoa, National Institute for Cancer Research, Genoa, Italy

This forum report contains conclusions about 3 different issues relevant to conducting clinical trials in deep mycoses. (1) Trials of diagnostic tests for deep mycoses must define the population appropriate for testing and the clinical question being asked. The unanswered question for the serum *Aspergillus* galactomannan assay is whether knowledge of results can change use of empirical therapy to treat febrile patients at high risk of invasive aspergillosis. (2) Use of historical controls is suboptimal but offers a pragmatic solution for studying rare mycoses; use of contemporaneous controls, matched for critical variables and evaluated by a blinded data review committee using detailed criteria, appears optimal. (3) Established groups of independent investigators, such as the European Organization for Research on Treatment of Cancer's Invasive Fungal Infections Group and National Institute of Allergy and Infectious Diseases's Bacteriology and Mycology Study Group, provide a pool of experienced investigators, defined operating rules, impartiality, and specialized expertise. Considering the enormous investment required for adequately powered efficacy trials of antifungal agents and the importance of these trials to guide clinical practice, use of collaborative groups outweighs the extra administrative time that is sometimes required.

ISSUES IN TRIALS OF DIAGNOSTIC TESTS

The sandwich ELISA (Platelia; BioRad) for detection of *Aspergillus* galactomannan antigen in patient serum was taken as a case study for the discussion on the design of diagnostic trials because of its clinical relevance to the other topics under discussion at the John E. Bennett Forum on Deep Mycoses Study Design. The members of the Forum used the studies by Maertens et al. [1] and Herbrecht et al. [2] as models for the discussion.

General principles for analysis of diagnostic tests. Four numbers encompass the simplest results of a diagnostic test. A tally is made of the patients with or without the diagnosis who have a positive or negative test result. Sensitivity and specificity use the 4 numbers to pose the question in a manner that assumes that the

[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):S123–7

1058-4838/2003/3608S3-0005$15.00

diagnosis is already known and asks the percentage of patients with the diagnosis for whom the test result was correctly positive (sensitivity) and the percentage without that diagnosis for whom the test result was correctly negative (specificity). Positive and negative predictive values look at the same 4 numbers but assume that the test result is known and ask whether the test correctly predicted the diagnosis. Both kinds of analysis require that an intermediate or indeterminate result is either excluded or reclassified as positive or negative. Both analyses assume that the cutoff separating categories has already been decided. Often the cutoff is not obvious and is decided as a compromise between sensitivity and specificity. A receiver operating curve (ROC) can help decide the optimal cutoff. The ROC plots rates of true-positive (sensitivity) versus false-positive (1– specificity) results at different cutoffs. The clinical implications of false-negative and false-positive results weigh heavily in estimating the impact of different cutoffs. The appearance of the ROC and choice of optimal cutoff will vary with the incidence of disease in the population under study (pretest probability). The principles elucidated above will be applied to the serum *Aspergillus* galactomannan ELISA.

Clinical results of galactomannan testing. The galactomannan ELISA results are reported as a ratio between the optical density of the patient's sample and that of a control with a low but detectable amount of galactomannan. The manufacturer considers a ratio >1.5 to be positive, a ratio of 1.0–1.5 intermediate, and a ratio <1.0 negative. The reported results with the galactomannan ELISA have depended on several variables: the degree of severity of the patients' aspergillosis at time of diagnosis, the prevalence of aspergillosis among the patients studied, the cutoff ratio used, and whether two consecutive positive tests were required for significance. Also, age has been important because more false-positive results have been found among young children [2]. In a patient population with a high incidence of aspergillosis, 71 autopsy-proven cases, a cutoff ratio of 1.5, and a requirement for 2 consecutive positive tests, the test had a sensitivity of almost 93% and specificity of 95% [1]. In a study of adults and children with less definite and less fatal disease, a lower prevalence of aspergillosis, and a single test with a cutoff ratio of 1.5, sensitivity was 28% (40/145 episodes) and specificity was 99% [2]). As these examples show, performance of the galactomannan test has depended critically on the population under study and appears most appropriate for patients with a high pretest probability of having the diagnosis.

Posing the clinical questions for future analyses. Another way to examine this diagnostic test is to identify the clinical questions that the test seeks to answer. In a patient with a high pretest probability of disease, such as an allogeneic hematopoietic stem cell transplant recipient with pulmonary infiltrate not responding to current antifungal therapy, the clinical question is whether to switch to, or add, another antifungal agent that would be useful to treat aspergillosis but not other molds. The clinician would definitely not make this switch if the test results were negative and the test had a high negative predictive value, because a negative test would mean that aspergillosis was very unlikely. If the test results were positive and the test had a high positive predictive value, the clinician might feel justified in switching to or adding an echinocandin to the regimen because the odds were high that the patient had aspergillosis.

Patients with an intermediate pretest probability of disease might include allogeneic bone marrow transplant recipients or relapsed leukemic patients with new fever, recently initiated antibacterial therapy, and chest CT that showed no pulmonary infiltrate. The clinical question in this case is whether empirical antifungal therapy should be started without waiting for a response to antibacterial therapy. A negative test result could not be used to decide whether to start or stop empirical antifungal therapy, irrespective of the positive and negative predictive value, because the test detects only aspergillosis and not candidiasis or other deep mycoses. On the other hand, a positive test result would support a clinical decision to start antifungal therapy with an agent that includes activity against *Aspergillus,* but only if the positive predictive value were very high, that is, the odds highly favored that all positive test results meant aspergillosis. Otherwise, too many patients would receive unnecessary antifungal therapy.

In the case of patients with a low pretest probability of disease, such as a neutropenic, recently febrile autologous hematopoietic stem cell transplant recipient, the clinical question might be whether a positive test can be used as a reason to simply repeat the test, to admit the patient to the hospital, or to obtain a high-resolution CT. The clinician could use a positive test result to make this decision if the negative predictive value were high, even if the positive predictive value were low, because the consequences of a false-positive result are modest. At worst, a false-positive result would lead to unnecessary radiological studies or hospital admission. A high negative predictive value is highly desirable in this clinical setting because clinicians have to trust a negative test result to correctly exclude patients who have early invasive aspergillosis.

Variability of results. Many variables determine how the *Aspergillus* galactomannan test performs. Verweij et al. [3] found 25% variability in assay titers between the 6 hospitals in their study. This level of variability is a significant concern if the test is going to be widely used as a part of the standard diagnostic workup in multicenter clinical studies. Batch-by-batch variability within individual centers has been reported but appears to be less prevalent than variability between centers. The use of prior triazole prophylaxis is an additional variable to consider, because this has been shown to delay the time to antigen positivity in both animal and clinical models [4].

Sequential testing. Sequential galactomannan tests that show repeated positive results, particularly with a rising titer, have been advocated to increase the positive predictive value. The time required to obtain additional tests inserts an unwelcome delay into the clinical decision-making process. Published studies do not make it clear how long one must wait to repeat a test after obtaining a positive result but do indicate that a new specimen is required. Repeating a test that yielded a positive result on the same specimen will usually produce the same result.

Analysis of the published *Aspergillus* galactomannan ELISA data indicates that no single interpretation applies to all patients. For patients at high risk of having invasive aspergillosis, obtaining 2 consecutive positive test results offers strong support for the diagnosis. The test appears insensitive in early disease and may not affect the use of empirical therapy. The test may prove useful in antifungal trials as a criterion for probable invasive aspergillosis, as has been proposed [5], and should be evaluated in that setting.

USE OF HISTORICAL CONTROLS IN ANTIFUNGAL THERAPY TRIALS

Use of historical control data for registration of antifungal agents has been common in the past. For example, approvals of azole drugs for treatment of blastomycosis, histoplasmosis, and coccidioidomycosis were based on open-label, dose-finding studies that compared results with published studies of amphotericin B or no treatment. The lipid formulations of amphotericin B were approved for the treatment of aspergillosis in patients who were intolerant of approved drugs or whose disease was refractory to prior therapy, and results were compared with historical controls. However, conclusions based on historical control data are known to be fallible in assessing treatment efficacy because of the confounding effects of poor data quality, variation in patient selection, and outcome evaluation bias [6, 7]. In addition, unmeasured confounders are problematic. The Forum endorsed the criteria proposed by Rex et al. [6] for good historical controls: clear and detailed inclusion and exclusion criteria, scoring of outcomes by a panel of blinded reviewers, use of objective and consistent definitions of outcome appropriate for the patient population, use of contemporaneous patients, and analysis using both a cohort design and a matched case-control design. Additional approaches that further enhance the ability to interpret the results of historical controls include distinguishing between enrollment in a salvage trial for clinical failure versus for drug intolerance, prospective collection of adequate data by use of better-designed case report forms, and the availability of large and detailed reference databases from previous trials. Establishing databases with results of prior trials would facilitate use of historical controls. However, the 2 major impediments to this approach are changes in medical practice over time and the commercial reality that the results of each trial belong to the company that funded it.

Historical controls in trials of therapy for invasive aspergillosis. The concerns about historical controls are particularly pertinent in light of the fact that a new echinocandin, caspofungin, was recently licensed by the US Food and Drug Administration (FDA) solely on the basis of a comparison of patients receiving caspofungin salvage treatment for aspergillosis and contemporaneous historical controls given amphotericin B. Despite FDA approval, some clinicians have expressed a reluctance to accept that caspofungin has demonstrated sufficient clinical efficacy in treating aspergillosis, given that the noncomparative study involved only 63 cases. In addition, the FDA advisory committee had some reservations about the manner in which the caspofungin data had been analyzed to draw detailed conclusions about the efficacy of the drug, rather than simply to observe the differences in outcomes between the caspofungin salvage group and the amphotericin B historical controls. The consensus at the Forum was that a randomized, double-blinded comparison should be the goal in the future and that acceptance of historical controls to license a drug sets an unwelcome precedent.

Historical controls in trials of therapy for endemic mycoses. Unlike the population with invasive aspergillosis, non-AIDS patients with selected endemic mycoses—for example, blastomycosis, histoplasmosis, and coccidioidomycosis—differ little from those of 50 years ago. Case definitions have been relatively standardized. Diagnosis of these diseases is relatively straightforward in comparison with aspergillosis, and there is little need for the development of new surrogate markers. Therefore, the use of historical controls may still be a valid approach for certain populations of patients entered into trials of new antifungal agents for endemic mycoses.

THE ROLE OF COOPERATIVE GROUPS IN THE MANAGEMENT OF ANTIFUNGAL CLINICAL TRIALS

Study groups dedicated to the improvement of trial design and interdisciplinary cooperation have advanced the management of antifungal clinical trials significantly in recent years. The groups that perhaps have made the greatest contributions to these advances include the Mycoses Study Group of the National Institutes of Health and the Invasive Fungal Infections Group (IFIG) of the European Organization for Research on Treatment of Cancer (EORTC).

The Mycoses Study Group was established in 1978 under the auspices of the National Institute of Allergy and Infectious Diseases. The Bacteriology and Mycoses Study Group (BAMSG), as is it now known, consists of experienced inves-

tigators at >50 academic medical centers across the United States. BAMSG is the prototype of a group that has successfully bridged the gap between academia, federal government, and industry. It has initiated 52 large-scale multicenter trials and completed 36 of these trials.

IFIG, established in 1991, has conducted a number of large-scale clinical studies of new antifungal agents, including the voriconazole and liposomal amphotericin B (AmBisome; Fujisawa Healthcare) phase III invasive aspergillosis studies. In addition to IFIG, the EORTC has an active Antimicrobial Therapy Group, which has led the way in the development of trials of empirical therapy for the management of febrile neutropenic patients since its inception in 1975. In the future, it is possible that the 2 complementary EORTC groups may merge, but in the meantime, both groups are looking for further opportunities to collaborate, both within the EORTC and externally with groups such as BAMSG.

Disadvantages. Although the advantages of cooperative groups undoubtedly outweigh the disadvantages, there are a number of limitations. Increasingly complex regulatory issues about patient-oriented research necessitate close scrutiny by all involved parties concerning study design, content of informed consent, and institutional review board documents. Consequently, the layers of administrative processing of concepts and protocols can result, at times, in slow initiation of BAMSG studies. This process may be a barrier to cooperation with pharmaceutical companies, which must find the elusive compromise between running a high-quality study and completing it in a timely and cost-effective manner.

Data sharing between industry and clinical trial groups can be a source of conflict. Because the pharmaceutical industry often files the Investigational New Drug Application with the US FDA and European regulatory agencies, the responsibility for collecting, verifying, and maintaining the confidentiality of the case reports for many studies lies with the pharmaceutical company sponsoring the drug. The reporting of safety problems and final analysis of the data in the New Drug Application are also the responsibility of the pharmaceutical company. The clinical trial group needs access to sufficient data to make their own analysis. The process and depth of data sharing vary with the study but require the pharmaceutical company to provide time, expense, and willingness to cooperate. Without mutual trust and effort on both sides, the clinical trial group may be asked simply to endorse data analysis provided by industry.

Advantages. Cooperative groups, such as EORTC and BAMSG, offer several major advantages to the evaluation of new drugs. In particular, each group offers a large pool of investigators and study centers with validated experience and productivity. These groups also have the facilities to run large, statistically well-powered studies. Many recently published antifungal studies have had insufficient power to be of any clinical value. Such studies represent an inefficient use of valuable resources. BAMSG and IFIG also offer specialized expertise in clinical and laboratory mycology, particularly in the areas of study design and independent data analysis and interpretation. They also play a pivotal role in the training and development of new clinical researchers.

Maintaining neutrality and impartiality is a key to the success of cooperative groups. The structure and open financial arrangements of BAMSG and IFIG, as well as the open discussions they initiate, ensure that personal biases and personal aggrandizement are minimized.

Cooperative groups can be instrumental in helping to move a study forward for publication. Authorship disputes are an increasingly common problem in multicenter trials. An established procedure, agreed upon before a study, for recognizing individual contributions can minimize such problems in a study group. The presence of writing committees that offer peer review further enhances the manuscript development process.

Cooperative groups such as EORTC and BAMSG can add value to industry trials of new antifungals by initiating concepts or suggesting improvements in trial design, offering an independent analysis of drugs across the antifungal field, providing a strategic, long-term view of needs within the field, and bringing to industry forward-thinking concepts for new trials. Cutting-edge trials should be the top priority for large clinical trial groups. Trial groups within the EORTC other than IFIG can be a valuable resource to the IFIG by drawing on their experience with other classes of drugs. However, these benefits to antifungal trials depend critically on successful interactions between the investigators, the pharmaceutical industry, and regulatory agencies. Discussion must begin early during trial design for the process to benefit from the trial group's expertise. Frequent investigator meetings can speed trial development, solve problems during study execution, and create harmony. Furthermore, maintaining morale among the clinical research nurses and coordinators at individual sites through regular communication of study progress is vital to the overall success of the venture.

Clinical trial groups should consider taking on the broader mandate of holding workshops on study design or construction of guidelines. Representatives from industry, regulatory agencies, and the practice community can add important perspectives. Better lines of communication and better antifungal drugs are a shared goal.

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 Develop-

ment, 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), and Tania Sorrell (Westmead Hospital, Westmead, Australia).

References

1. Maertens J, Verhaegen J, Demuynck H, et al. Autopsy-controlled prospective evaluation of serial screening for circulating galactomannan by a sandwich enzyme-linked immunosorbent assay for hematological patients at risk for invasive aspergillosis. J Clin Microbiol **1999**; 37:3323–8.
2. Herbrecht R, Letscher Bru V, Oprea C, et al. *Aspergillus* galactomannan detection in the diagnosis of invasive aspergillosis in cancer patients. J Clin Oncol **2002**; 20:1898–906.
3. Verweij PE, Erjavec Z, Sluiters W, et al. Detection of antigen in sera of patients with invasive aspergillosis: intra- and interlaboratory reproducibility. The Dutch Interuniversity Working Party for Invasive Mycoses. J Clin Microbiol **1998**; 36:1612–6.
4. Petraitiene R, Petraitis V, Groll AH, et al. Antifungal activity and pharmacokinetics of posaconazole (SCH 56592) in treatment and prevention of experimental invasive pulmonary aspergillosis: correlation with galactomannan antigenemia. Antimicrob Agents Chemother **2001**; 45: 857–69.
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. Rex JH, Walsh TJ, Nettleman M, et al. Need for alternative trial designs and evaluation strategies for therapeutic studies of invasive mycoses. Clin Infect Dis **2001**; 33:95–106.
7. Powers J. Counterpoint: alternative trial designs for antifungal drugs—time to talk. Clin Infect Dis **2001**; 33:107–9.