Cohorts, Trials, and Evidence: Expanding Our Confidence in Guidelines for Antiretroviral Resistance Testing

Practice sometimes gets ahead of the evidence. Physicians frequently base decisions on evidence with less strength than the “gold standard” of prospective randomized trials with clinical end points. Care guidelines may advocate such decisions on the basis of preliminary research, expert opinion, and data derived from clinical cohorts. Reliance on imperfect evidence is not necessarily a problem, particularly if the intervention is plausible given the current understanding of disease pathogenesis (face validity). Recommendations based on such evidence are more likely when treatments or diagnostic technology advance so rapidly and become so widely accepted that the equipoise needed for randomized trials is lost. In this case, trials simply are not feasible. In this issue, Palella and colleagues (1) address an example of this problem.

HIV can become drug resistant when it is allowed to replicate in the presence of antiretroviral agents (2). Drug resistance can limit treatment benefit and often requires a change in the prescribed regimen (3). Drug-resistant retroviruses can also be transmitted (4), potentially compromising the host’s response to drugs to which HIV is already resistant. Awareness of these facts emerged shortly after antiretroviral drugs were developed and the technology for detecting drug resistance became widely available (5). The 2 types of HIV drug resistance assays are genotyping and phenotyping. Genotyping, the faster and less expensive method, reports genetic mutations known to be associated with resistance, whereas phenotyping tests the ability of the virus to grow in the presence of specific antiretroviral drugs (6). These tests rapidly penetrated the marketplace of HIV care. They were accurate; their conceptual foundation closely resembled bacterial resistance testing, which was already widely accepted; and testing was plausible within the framework of the biology of HIV infection. Whether routine testing for drug resistance improved mortality, however, was never rigorously tested in randomized trials. Despite this shortcoming, many guidelines recognize HIV drug resistance testing to be a cost-effective standard of care (7, 8) at baseline in untreated persons to detect transmitted drug resistance and in cases of recurring viremia after initial suppression of chronic infection (9–11).

In a cohort study, Palella and colleagues (1) investigated the utility of HIV drug resistance testing by analyzing patient survival as a function of whether a physician had ordered either a genotype or phenotype resistance test (GPT). Their report is instructive and highlights the advantages and limitations of observational research versus randomized clinical trials. The HOPS (HIV Outpatient Study) is a Centers for Disease Control and Prevention–supported study in which data have been collected prospectively from 10 U.S. sites since 1993. Mostly cared for in university or private practice clinics, participants are disproportionately white, male, engaged in homosexual behavior, and privately insured. Palella and colleagues identified 3 subgroups: patients naive to antiretroviral therapy; those currently receiving therapy; and those who had received all classes of antiretroviral drugs available from January 1999 to December 2005, but in whom therapy failed. They measured the use of GPT, as evidenced by the medical record, and evaluated its association with mortality in the overall cohort and in the 3 subgroups. After adjusting statistically for baseline between-group differences in factors that were expected to contribute to lower mortality (including higher CD4 cell counts, white race, private insurance, and no injection drug use), they found that GPT was associated with an overall survival advantage. The advantage was most evident in the subgroup receiving antiretroviral therapy at baseline. Testing did not have a significant benefit in those who were treatment-naïve at baseline. Palella and colleagues did not compare the actual results of GPT or any resulting adjustment in treatment regimens, an analysis that the authors stated might be instructive. They conclude that ordering GPT in HIV care is associated with decreased mortality, particularly in patients already receiving treatment at the time of testing.

The results are encouraging and reinforce current guidelines for drug resistance testing in HIV care. Their article also resonates with an ongoing debate about the strength of “evidence” from observational studies, in which receiving an intervention reflects the clinical circumstances of the individual patient, versus evidence from prospective clinical trials, in which the intervention is randomly assigned. The debate about the weight that observational studies should get in the development of treatment guidelines is unresolved. The lack of apparent benefit of GPT in the treatment-naïve subgroup may reflect an inadequate sample size but may also result from the current use of much more potent regimens than those previously used. Virologic failure with these regimens is uncommon. When it occurs, medication adherence is often so poor that exposure to the drug is too limited to generate drug resistance. Reduced benefit of GPT in very advanced drug resistance, which occurred in the study, is also not surprising. In the time frame of their study, successful “salvage therapy” of patients with virologic failure was rare, principally because the powerful drug classes now in use were not available. Evidence, whether gained from clinical trials or cohorts, is always evolving, and guidelines should be subject to revision.

The authors discuss the central challenge of cohort analyses such as theirs, which is unmeasured confounding: That is, an observed effect is due not to its apparent cause (for example, GPT) but to an unmeasured factor that af-
fects the decision to test and the outcome of the intervention. Although Palella and colleagues used contemporary statistical techniques to adjust and analyze their data, concern remains that the 3 study groups and overall cohort may differ enough in characteristics that affect the decision to order GPT and measured outcome of HIV infection to account for the observed differences. A bias of this type does not occur in well-conducted, randomized clinical trials, and the choice between the 2 means of collecting clinical evidence is a subject of intense debate in HIV care. Recently, for example, researchers analyzed 2 large HIV composite cohorts (12, 13) to determine the mortality benefit of initiating antiretroviral therapy at various CD4 cell count thresholds. In 1 cohort (12), they found reduced mortality only at CD4 counts less than 0.350 cells × 10^9 cells/L. In the other cohort (13), the researchers found a significant survival advantage even when the count was greater than 0.500 × 10^9 cells/L. Both sets of investigators spent considerable effort making the different treatment threshold cohorts as similar as possible, and yet their conclusions were quite different.

Some argue that observational data are sufficiently reliable to be used in forming treatment guidelines, particularly after sensitivity analysis is done (to see whether the conclusions change when a hypothesized confounder is assumed to be present) and similar techniques are used to test for unmeasured confounding and other biases. Others argue just as forcefully that the limitations of cohort studies still require us to conduct definitive prospective controlled trials. In both cases discussed here (when to order GPT and when to initiate antiretroviral therapy), there may be no certain pathway to resolve these debates. Few physicians challenge the use of GPT in routine clinical care, few patients would be sufficiently close to equipoise to enroll in a trial limiting its availability, and human subjects protection committees would probably consider such a trial to be unethical. Conducting a definitive prospective trial comparing optimum times to initiate antiretroviral therapy is also a daunting task. The small expected differences in actual mortality given the potency of established antiretroviral drugs and recent cohort results would demand a very large and very long trial that almost necessarily enrolls patients who vary greatly in HIV type, comorbid diseases, and other factors that might limit the generalizability of the findings. Changes in the specific drugs prescribed, which is predictable in the field of HIV therapy, would be another problem with a large, long trial.

In the end, those who develop clinical policy must often cope with imperfect evidence. One framework for thinking about evidence is that its function is to change the probability that a conclusion is correct. Randomized trials often, but not always, move the probability closer to 1.0 or 0 than observational studies will, but we do not need perfect information to act. In this decision-making framework, although cohort analyses have inherent limitations, Palella and colleagues’ study and studies performed to address the threshold CD4 count for treatment initiation are important. Evidence generated by observational studies is more credible than expert opinion and deserves to be taken seriously, particularly when controlled trials are impractical.

Paul Volberding, MD
San Francisco Veterans Affairs Medical Center and University of California, San Francisco
San Francisco, CA 94121

Potential Financial Conflicts of Interest: Consultancies: Bristol-Myers Squibb, Pfizer, Merck, Gilead, Schering, GlaxoSmithKline, TaiMed.

Requests for Single Reprints: Paul Volberding, MD, Veterans Affairs Medical Center, 4150 Clement Street, San Francisco, CA 94121; e-mail, paul.volberding@med.va.gov.


References