

cine with the goal of simply becoming competent. I want to be the kind of doctor whom a patient can trust, the kind who listens and touches and takes the

time to look at a patient to see how he or she is truly faring. For now, I can count my improved sense of smell as a small victory in that quest.

(Identifying details about the patient have been changed to protect his privacy.)

Dr. Bomback is a resident in internal medicine, University of North Carolina Hospitals, Chapel Hill.

## FOCUS ON RESEARCH

# First, Gather the Data

David Hunter, M.B., B.S.

Related article, page 353

It is a cornerstone of medical practice to “first, do no harm.” Yet the body of evidence that is sufficient to demonstrate efficacy for a new drug is rarely large enough to provide absolute assurance that harmful side effects do not exist. Thus, ongoing surveillance is necessary to detect adverse events.

There are many reasons why the randomized trials that are adequate for demonstrating drug efficacy may not be adequate for the recognition of important side effects (see table). Randomized trials may be too small to permit the detection of adverse events. For instance, for a continuous outcome such as blood pressure or the change in a pain scale, in a trial with a before-and-after or crossover design, even a few dozen patients may be an adequate number to demonstrate an effect. Trials are often conducted over weeks or a few months, obviating any possibility of detecting longer-latency effects that require exposure for many months or years. Finally, trials may be conducted in a population from which patients with coexisting illnesses have been excluded and thus do not address the question of whether the drug may do harm once approved and prescribed to these patients.

The detection of adverse events also depends on the background incidence of these conditions. The occurrence of two cases of progressive multifocal leukoencephalopathy among 3000 patients treated with natalizumab (Tysabri) in clinical trials was enough to cause the drug to be withheld from the market. Less remarkable conditions, such as heart attacks associated with the use of cyclooxygenase-2 (COX-2) inhibitors, are less likely to raise flags. To detect an increase in the incidence of these conditions requires careful statistical analyses of unblinded data and adequate study power. In this paradoxical manner, we may be better protected from exposure

to drugs that cause very rare medical conditions than from exposure to those that cause common, unremarkable conditions.

Once concern has been expressed about a drug and specific adverse events, it is often proposed that the Food and Drug Administration mandate that companies conduct clinical trials to prove whether the drug does or does not cause the events. This approach is always preferred, but such studies may have to be almost unfeasibly large, and for drugs shown to be efficacious in the treatment of serious conditions, it may be argued that it is unethical to conduct a placebo-controlled trial to search for adverse events (see table). In

### Some of the Reasons That Results of Randomized, Controlled Trials May Not Be Available for Assessment of Adverse Events Associated with Prescription Drugs.

Trials powered for efficacy may be too small to detect adverse events.

Monitoring of adverse events may not be sensitive or specific for the actual events caused.

Duration of trials may be too short for detection of events requiring longer exposure.

Stopping rules in clinical trials may further shorten the duration of exposure after randomization.

Enrollment criteria may exclude susceptible subgroups.

For industry-sponsored trials, head-to-head comparison of adverse events due to drugs from different manufacturers may not be available.

Follow-up studies to detect adverse events that involve the denial of an efficacious medication to patients may be deemed unethical. Patients may not wish to enroll in such a study.

Funding to conduct trials solely to quantify adverse events may be difficult to obtain.

some cases, adverse events have come to light in trials that studied a drug's usefulness in preventing a condition other than the original indication. In the wake of the withdrawal of most of the COX-2 inhibitors, however, it may be difficult to persuade companies to sponsor or cooperate in this type of trial, which was helpful in clinching the case that these drugs caused harm. In any case, it is only a minority of drugs for which evidence may suggest that such large-scale trials for a different indication would be worthwhile.

The main alternative to evidence from randomized trials is evidence from observational studies, in which the occurrence of adverse events in patients who receive a specific drug is compared with this occurrence in patients who do not. The obvious concern in these nonrandomized studies is that the patients who receive a specific drug may have a different level of risk of the adverse events under study from that in the patients who receive no therapy or other drugs — that is, the drug–outcome association is confounded by demographic and lifestyle characteristics or by the presence of coexisting conditions. As in any observational study, there are a variety of design strategies (e.g., matching patients according to potential confounding factors) and methods of analysis (e.g., multivariate regression) that can be used to control for confounding. Yet it is not possible to control for confounding factors that have not been measured. Many drug-surveillance studies rely on large administrative databases, such as those of Medicare and

health maintenance organizations, in which information about individual lifestyle factors (such as cigarette smoking) may not be available, and information on coexisting illnesses or the adverse events themselves may have to be inferred from billing records or hospital admission or discharge codes. To the extent that the information is either nonexistent or only a weak surrogate for the actual events, even the most rigorous statistical methods cannot correct for confounding. Concern about the potential for residual confounding has led to reluctance to consider the results of these studies as anything more than hypothesis-generating or preliminary, and action that could modify the use of harmful drugs has been delayed as a result.

How, then, to reconcile the reasonable skepticism about the results of observational studies with the fact that they may be the only source of evidence on the side effects of drugs? In this issue of the *Journal*, Mangano and colleagues (pages 353–365) provide an example of the type of study that may be a model for the future. Faced with a question regarding the safety of specific antifibrinolytic agents given to minimize blood loss after certain cardiac surgical procedures, which they determined could not be answered with a randomized trial because of ingrained practice, the authors conducted a large multi-institutional study with the aim of optimizing the availability of data on potential confounding factors. The authors collected a large amount of data (about 7500 data fields per patient) on the characteristics of the patients, patients' co-

existing illnesses and selected adverse outcomes. This approach permitted the use of propensity-score and conventional multivariate techniques to control confounding and ensured that the clinical outcomes of interest were not obscured by misclassification due to a lack of clinical detail or an inaccurate administrative assignment of diagnostic codes. These prodigious data pointed to an increase in the risk of renal failure and cardiac events with the use of one of three drugs. Because extensive data were collected, the estimates of risk could be analyzed in conjunction with an extensive array of potentially confounding variables.

Pharmacoepidemiologists are exploring new ways to minimize the potential for confounding in observational studies of the effects of prescription drugs. For instance, the propensity-score approach estimates the probability that a person will be given a prescription for a particular drug on the basis of his or her demographic, lifestyle, and clinical characteristics; this score can then be used to control for potential confounding from these characteristics. Another potential application of the score is to match patients who received the study drug with control patients who did not but who have the same propensity score; in essence, this is an attempt to replicate the process of randomization, in which other unmeasured and potentially confounding characteristics are randomly distributed among those who receive a drug and those who do not. Obviously, a propensity score is only as good as the information used to

calculate it, and it does not account for unmeasured potential confounding factors — this accounting is the main virtue of randomization. Furthermore, in some analyses and simulation studies, the use of the propensity-score approach has not resulted in a measurable improvement in the control of confounding as compared with conventional multivariate methods involving the use of the same information.

In the final analysis, confounding cannot be controlled in an observational study unless information on the confounding factors, or on good surrogates for them, is collected for analysis. Studies such as the one by Mangano and colleagues point the way to the prospective design of studies to assess drug safety and to the collection of as much information as necessary to provide answers of the highest quality. Sub-

stantial questions remain about how to fund and administer these studies, but we need to ensure that skepticism about the value of observational studies does not engender nihilism. In the absence of evidence from randomized trials, the best-quality data must be made available to ensure the safety of medications.

---

Dr. Hunter is a professor of epidemiology at the Harvard School of Public Health, Boston, and a statistical consultant to the *Journal*.