Clinical Trial Monitoring: Hit or Miss?

A *Science* investigation reveals that there is no overall system for monitoring clinical trials in the United States. Some experts argue that there should be, particularly at NIH

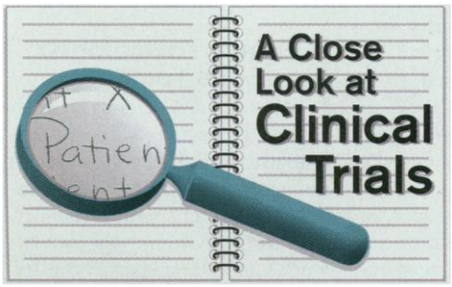
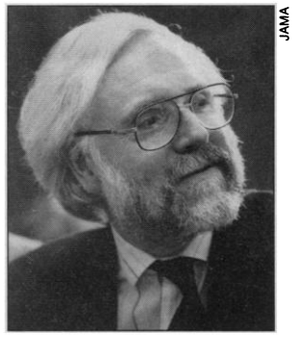


ILLUSTRATION: C. FABER SMITH

Large-scale clinical trials are crucial to contemporary medicine. These tests, which gather thousands of patients at many different clinical centers into large groups, provide the best—indeed often the only—means of assessing new drugs, vaccines, devices, and therapies. Yet because they are so large and rely on so many investigators, clinical trials are vulnerable to slipshod practices and even to the occasional instance of scientific fraud.

The National Cancer Institute (NCI) is currently receiving a painful lesson in just how damaging misconduct can be in a large clinical trial as it copes with publicity surrounding several studies that included tainted data from Roger Poisson of St. Luc Hospital in Montreal. Though NCI has no evidence that Poisson's data led researchers to false conclusions, the case shook many women with breast cancer and cast a pall over one study's conclusion that lumpectomy is as effective as mastectomy. Most experts think cases of misconduct as blatant as Poisson's are rare. (Among other things, he changed dates on patient records and enrolled a patient who had explicitly refused to participate.) But it's not easy to say how rare—or to say how common less egregious instances of slipshod practice are in clinical trials.



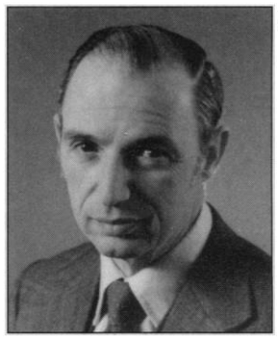
JAMA

Common sense. *JAMA*'s Drummond Rennie.

The reason these questions can't be answered with confidence is that there is no uniform and systematic oversight of clinical trials in the United States, as an extensive investigation by *Science* has revealed. This investigation, carried out over the past 3 months, shows that a patchwork of policies has developed over the years as agencies came up with (or had imposed on them) ad hoc solutions to problems in clinical trials. And, as the *Science* investigation shows, there are plenty of gaps and widely differing standards in this nonsystem. Examples include:

- No overall policy at the National Institutes of Health (NIH) regulates oversight of clinical trials.
- Some large, multicenter clinical trials sponsored by NIH do not include on-site audits of data.
- The U.S. Food and Drug Administration (FDA) does limited on-site auditing of data from "important" trials, but has no regulation stipulating that a study sponsor must conduct regular, on-site auditing of trial data.
- NCI requires audits of each site in a large trial once every 3 years. In contrast, the pharmaceutical industry has a much tighter system, with frequent audits of data at each trial site.

A growing number of critics contend that, unless there is an overhaul of clinical trial monitoring, there will be more revelations like the one involving Poisson. "A lot of us are giving our lives to helping men and women cure diseases, and we keep having to have people hit us on the head with two-by-fours," laments Drummond Rennie of the University of California at San Francisco, a deputy editor of the *Journal of the American Medical Association* (JAMA) who has strongly argued for oversight reform. "Why should we learn so slowly?"



FDA

Increased awareness. FDA's Alan Lisook.

In support of Rennie, the few studies that have been done on the subject find that sloppy practices short of misconduct occur frequently enough to be of real concern. And since many important clinical trials are too large and costly to repeat, some observers argue strongly for stepped-up monitoring.

Yet on-site auditing comes at a price. Even some experts in industry, which does rigorous monitoring, caution that intensive on-site auditing may be too expensive to be worthwhile—and could discourage future clinical research. Eve Slater, senior vice president for clinical and regulatory development at Merck Research Labs, asks, "Can you put into the system more checks and balances without sinking the ship? That's a tall order." The order could be particularly tall in today's cash-strapped climate at NIH, where every dollar spent on auditing might mean a dollar that is not spent on research.

NIH: No uniform policy

One of the most dramatic findings of the *Science* investigation (which focused mainly on "Phase III" clinical trials, the large studies that precede FDA approval of a drug or device) is that NIH has no overall policy governing oversight of large, multisite clinical trials. And since NIH currently sponsors more than 300 such trials, which shape health-care decisions for the entire country, the question of whether it should have an overall policy is a crucial one.

Though there aren't overall policies, or even guidelines, it should not be assumed that there are no safety mechanisms for NIH trials. On the contrary, many checks and balances are built into large, multisite clinical trials to weed out bad data or bad researchers. Typically, a university serves as a coordinating center and receives data, filed on a "case report form" tailored to the trial, from many dispersed medical centers. If one site's results look wildly different from the rest, data are missing, or discrepancies surface in records from one patient (as happened in the Poisson case), data managers often investigate. Independent Data Safety Monitoring Boards, which evaluate data in an ongoing trial to make sure patients aren't being harmed, provide another review. Institutional Review Boards (IRBs) at each medical center guard against ethical and protocol violations.

But the most rigorous check of data in a multisite study is auditing at the trial site: comparison of the trial's case report forms with original patient records. On-site auditing is the only type of monitoring intended to seek out sloppiness, carelessness, and fraud by comparing the patient's clinical record with the data entered in the trial.

Since on-site auditing is considered by

experts in the field to be the best safeguard against fraud and sloppiness in clinical trials, *Science* wanted to know how often that top-quality mechanism is used at NIH. No one at NIH could answer that question, since no one at NIH has overall responsibility for the conduct of clinical trials. Our own survey shows that while on-site auditing is done in many trials, it varies greatly in frequency and intensity—and in some trials it isn't done at all (see table on p. 1536).

Our survey began where the current scandal did—at NCI. The fact that the Poisson case happened there is ironic, since *Science* found that NCI has the most formalized on-site auditing program of any institute. Most large clinical trials sponsored by NCI are staged through 14 cooperative oncology groups (COGs), regional nets including 16,000 clinicians at 2200 institutions. In all, COGs run 135 large clinical trials—nearly half of the large trials sponsored by NIH. NCI policy requires that each site in a COG be audited at least once every 3 years, though some COGs are much less rigorous than others (see box on page 1537).

Outside the COGs, information for NCI is sketchy. In all, NCI is funding 58 multisite clinical trials through its standard grant system; another eight are run by NCI's Division of Cancer Prevention and Control. Michaele Christian, NCI's acting chief of the new Clinical Trials Monitoring Branch—established to increase oversight in response to the Poisson case (*Science*, 22 April, p. 499)—says she does not know whether auditing occurs at NCI-sponsored trials outside the COGs. The information is not available, Christian says, because there has never been a systematic evaluation of monitoring across NCI. Her branch is now re-evaluating NCI's entire oversight program.

The bulk of the large clinical trials sponsored by NIH that don't involve cancer focus on AIDS. NIH's largest network of clinical investigators, the AIDS Clinical Trials Group (ACTG) at the National Institute of Allergy and Infectious Diseases (NIAID), hires an outside contractor to monitor data. NIAID says the contractor audits about 50% of the records and visits each major site four times a year. An NIH insider familiar with auditing both at NCI's COGs and NIAID's ACTG says the contractor is "an order of magnitude" more thorough than the NCI review.

Unlike NCI and NIAID, the next-largest NIH sponsor of multisite clinical trials, the National Heart, Lung and Blood Institute (NHLBI), doesn't organize researchers into standing networks. Instead, each trial has its own network—and its own audit policy, which "varies from study to study," says Lawrence Friedman, director of NHLBI's division of epidemiology and clinical application. In terms of frequency and number of records audited, says Friedman, "the bottom line is we do not do what the drug companies do"—that is, conduct regular audits comparing original patient forms with the study's secondary records. For example, in the biggest NHLBI trial, an evaluation of digitalis in congestive heart failure, 7790 patients are enrolled at 303 sites. "We have no ability or intent to visit all of those sites," says Friedman, adding that a sample are visited.

Like NHLBI, the National Institute of Neurological Diseases and Stroke (NINDS), another big sponsor of multisite trials, doesn't "often do on-site reviews" that compare patient records to case report forms, says Michael Walker, director of NINDS's division of stroke and trauma. NINDS's largest study, the North American Symptomatic Carotid Endarterectomy Trial, including 2800 patients at 90 sites, has no on-site audits, Walker says.

The National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) recently completed two important multisite trials, one to see whether diet could reduce renal disease, the other to evaluate whether strict control of blood sugar could reduce diabetes complications. Walter Stolz, head of NIDDK's division of extramural activities, says neither study had on-site data audits. And Stolz says none of the five large trials on NIDDK's drawing board has such plans.

At the National Institute of Arthritis and Musculoskeletal and Skin Disease, the largest trial is designed to study whether therapy is more effective than amputation in treating severe trauma. This eight-site trial involving 700 patients has no required data auditing. "The possibility that data could be forged was not considered by us," says Michael Lockshin, director of the institute's extramural program. "All of us have undergone a paradigm shift in our thinking about the way trials should be done in response to this [breast-cancer fraud] crisis."

Industry: More frequent auditing

In contrast to NIH, frequent on-site auditing of multisite clinical trials is the norm for pharmaceutical companies. Those audits aren't required by the FDA, says Alan Lisook, head of the FDA's clinical investigations branch. In fact, the agency has no rule requiring on-site auditing of large clinical trials. In 1977, the agency tried to introduce rules requiring annual auditing, but Lisook says they never became law because of "a change in Administrations." Yet the proposed rules did have an effect: The pharmaceutical industry started to perform frequent audits in part because of them, says Lisook. And in 1988, FDA issued guidelines encouraging sponsors to have monitors "frequently" visit trial sites and make sure investigators follow study protocols, keep accurate records, and report to their IRBs.

Pharmaceutical companies, anxious for FDA approval of their products, tend to comply with these voluntary guidelines. At Burroughs Wellcome Co., for example, Dick Kiernan, director of international regulatory compliance, says his company first sends monitors to select sites and evaluate whether a clinician is qualified. Once the study begins, monitors frequently return to each site—every 6 to 10 weeks, say—to check records and make sure the site is following the protocol. In addition, auditors working for the company make surprise visits and check the monitors. "If there are problems, I don't want the FDA to find them," says Kiernan. "I want my compliance people to find them so they can correct them. There's nothing worse than to have an FDA inspector say he's found a problem and a clinician to say, 'Oh.'" Merck has a similar program, explains VP Eve Slater.

Merck and Burroughs Wellcome, two major pharmaceutical companies, illustrate how industry is anxious to beat FDA to the punch. But by and large, NIH has a different mindset. NIH officials do acknowledge that the Poisson case has shaken them. "It's very unfortunate when people begin to disbelieve major findings of a study," says NIH director Harold Varmus, referring to the lumpectomy trial. "Very important issues have been raised

TRIAL TERMINOLOGY

Data Safety Monitoring Board (DSMB)—Researchers, ideally independent of a clinical trial, who periodically review data in blinded, placebo-controlled trials. DSMBs can prematurely stop a trial either if toxicities are found or if treatment is proved beneficial.

Institutional Review Board (IRB)—Researchers, ethicists, and lay people who safeguard trial volunteers. Primarily review trial protocols and informed consent forms

Coordinating center—Headquarters for a multisite trial that collects all data

Case report forms—Standard documents used by clinicians to report patient data to coordinating center

Patient records—Patient charts used at hospitals, clinics, and physicians' offices

On-site auditing—Comparing original patient records against case report forms

Monitoring—Used interchangeably with auditing, but also refers to checking data sets for anomalies

here, and some of the practices are not to be tolerated." Yet Varmus isn't sure NIH should move toward the industry model. "We do give more autonomy to grantees," he says, and "we need to titrate oversight versus the need to encourage independence."

Researchers' independence isn't the only issue that makes top NIH officials think twice before blithely instituting a policy of mandated on-site auditing for their clinical trials. Another—and a crucial one—is cost. NCI's COGs spend about 0.1% of their $750-million budget, or $750,000 a year, on auditing, a figure NCI officials say is less than the actual price of the audits because some personnel charges are accounted for separately. Last year, NIAID paid its contractor, Pharmaceutical Products Development Inc., $3.48 million to oversee 57 ACTG sites and 11 others, a figure that includes training for auditing personnel as well as the cost of auditing. Augmenting clinical-trial monitoring at NIH would mean paying even more—at a time when NIH is already being forced to make painful cuts. "How do we do more monitoring and oversight when we're told to decrease our personnel in response to [President] Clinton's mandate to reduce the size of government?" asks Varmus. "It's a stretch."

Since the costs of on-site auditing are so tangible, particularly in the current fiscal environment at NIH, the benefits would have to be just as clear to make mandatory auditing worthwhile. Are they? One instructive study was reported by an NCI-sponsored COG in *JAMA* last July. The study reports the results of 11 years of auditing one COG, the Cancer and Leukemia Group B. More than 200 sites now participate in this COG, and each is audited at least once every 3 years. The auditing has uncovered one case of fraudulent data and another of "gross scientific error," suggesting to the paper's authors that scientific misconduct is "very rare" (0.28% of the audits done).

Yet the study did find "many errors of omission and commission" that were partly responsible for the COG dropping three major medical institutions and 96 affiliate sites. For example, the auditing uncovered "major protocol deviations" in drug dosing in more than 10% of the patient records; when the auditing first started, 49.6% of the affiliates had major deviations in their IRBs (some didn't even have an IRB), and 10% of the patients enrolled weren't eligible. Raymond Weiss, chief of medical oncology at the Walter Reed Army Medical Center and the study's first author, says that because his COG weeded out bad sites and created an atmosphere where the remainder felt pressure to monitor themselves, nearly every parameter the auditors evaluate has improved over time. "People are well aware that they have others looking over their shoulders," says Weiss, who heads the COG's audit committee.

Some other information on the value of close monitoring comes from the FDA, which also tracks the results of its audits. Those audits spot-check a percentage of important trials that might lead to a new drug's approval. Between June 1977 and January 1994, FDA conducted 3092 on-site inspections and found that 56% of sites had problems with patient consent forms, 22% could not adequately account for the drugs they were dispensing, 29% did not strictly adhere to the protocol, 23% had inadequate and inaccurate records, 12% had IRB problems, and 3% had a significant fraction of records missing. Lisook says that since 1977, "truly egregious problems," including those that lead the FDA to bar researchers from doing clinical trials, dropped from 11% of trials to 5%. "It appears that things have been getting better with increased awareness and with increased oversight," he says.

Results like these lead critics of current clinical-trial practice to argue that regular on-site auditing is a necessity, not a luxury. "People in science trust each other and say, 'This is something we don't need to spend money on,'" says Weiss. "But I know there are no perfect patients, no perfect doctors, and no perfect data managers." *JAMA*'s Drummond Rennie believes on-site audits are "commonsensical" in large clinical trials.

But for that view to become official at NIH, the current scandal must have a broader effect than previous ones—each of which has led to ad hoc fixes in the monitoring system. In 1980, a time when NCI's

NIH: A PATCHWORK OF MONITORING STRATEGIES

Institute	Clinical Trial	Sites/Subjects*	Auditing
Aging	Osteoporosis prevention/intervention	4/1121	No on-site
Allergy and Infectious Diseases	AZT in healthy HIV-infected	31/3236	Every site, 4x/year, about 34% of records
	Combo drugs in HIV-infected	43/2495	Every site, 4x/year, about 78% of records
	Acellular pertussis vaccine in Sweden	2450/48,623	10–15% of sites/year, 100% of records
Arthritis and Musculoskeletal and Skin Diseases	Amputations vs. treatment in severe trauma	8/700	No on-site
Cancer	Adriamycin for breast cancer	240/2400	Every site, 1x/3 years
	Tamoxifen for breast cancer	240/2160	Every site, 1x/3 years
	Wilms' tumor therapy	120/1705	Every site, 1x/3 years, 5–10% of records
Child Health and Human Development	Neonatal intensive care	12/38630	1–2 sites/year, 5–10% of records
	Maternal fetal medicine	11/9626	1–2 sites/year, 5–10% of records
Diabetes and Digestive and Kidney Diseases	Therapy of benign prostatic hyperplasia	16/2472*	Not routine
Environmental Health	Treatment of lead-exposed children	4+/1332	Every site, 1x/year
Eye	Age-related eye disease	13/4600	Every site, 1x/year, 5–10% of records
	Collaborative ocular melanoma	50/3000	Every site, 1x/3 years, >10% of records
Heart, Lung and Blood	Digitalis in congestive heart failure	303/7790	Sample of sites visited, infrequent
	Bypass angioplasty revascularization	15/1829	Now in follow-up; every site visited 2x
Neurological Diseases and Stroke	Symptomatic carotid endarterectomy	90/2800	No on-site
	Tirilazed in stroke	50/2260	No on-site
	ORG-1072 in acute ischemic stroke	35/1800	Organon Inc. does all sites 1x/6 weeks

* Sites may include coordinating centers and other core support centers. Some subject numbers are projected enrollments.

The Poisson Case: Battle Over Auditing

Ever since the *Chicago Tribune* reported on 13 March that fraudulent data had tainted National Cancer Institute (NCI) studies of lumpectomy versus radical mastectomy for treatment of early breast cancer, much criticism has been directed at Roger Poisson, the researcher at St. Luc Hospital in Montreal who falsified and fabricated data for 90 patients. The cooperative oncology group (COG) that ran the trials, the 35-year-old National Surgical Adjuvant Project for Breast and Bowel Cancers (NSABP), has also been pilloried, along with NSABP chair Bernard Fisher of the University of Pittsburgh, for not reporting the bad data earlier.

But with most of the media focused on the drama of personalities and organizations, little attention has been paid to just how inadequate the data monitoring mechanisms in these trials were. As shown in correspondence obtained by *Science*, the Poisson case made NCI officials aware 2 years ago that the study's data monitoring didn't meet current standards. But when NCI officials tried to get Fisher and his colleagues to improve their monitoring system, they were rebuffed.

In 1992, NSABP policy was to oversee trials through the group's biostatistical center in Pittsburgh, which compared report forms submitted by physicians with photocopies of patients' primary records. In addition, NSABP did on-site audits of a sampling of patient records—typically eight—at each of its 292 sites. On 13 July 1992, NCI officials wrote Fisher that because of the St. Luc case, triggered when the biostatistical center discovered two operation reports with different dates for the same patient, NCI had reviewed NSABP auditing procedures and found them "obsolete." "We strongly urge NSABP to review their audit procedures," wrote Richard Ungerleider and Joan Mauer, officials in NCI's Division of Cancer Treatment. They also suggested NSABP follow the lead of other NCI-funded COGs and conduct more extensive on-site audits.

On 1 September, Fisher and Carol Redmond, director of NSABP's biostatistical center, fired back a letter saying NCI's "statements could be interpreted as indicating that some negative association exists between the putatively obsolete NSABP audit program and the investigation of an NSABP institution....[W]e not only reject such a consideration but find it both perplexing and paradoxical that criticism is being directed toward the very mechanisms that enabled us to identify, report, and thoroughly investigate the discrepancies." After reviewing audit procedures of other COGs, Fisher and Redmond say they concluded that "the audit program of the NSABP is at least as good as, if not better in many respects than, that of other groups." (Fisher declined to be interviewed to discuss this letter, and Redmond did not return repeated phone calls.)

In a letter the following month, NCI's Ungerleider and Mauer pointed out that St. Luc had first submitted false reports to NSABP 13 years previously—and that therefore "one might... question whether the audit policy of reviewing only eight patient records at an institution that entered over 400 patients in a 3-year period contributed to the delay in detecting significant data irregularities."

To make matters worse, in April NCI announced that it had done on-site audits at NSABP sites at Louisiana State University (LSU) and Tulane University and found that the "standards of record keeping and reporting at these institutions were unacceptable." At Tulane, auditors were able to confirm eligibility by the study's criteria for fewer than 20% of patients enrolled in NSABP studies; the LSU audit could confirm eligibility in only 40% of patients. Acceptable informed consent could be found for only 75% of the LSU patients.

Because of "deficiencies" in auditing and quality assurance, NCI has removed Fisher as NSABP's chair, placed the group on probation for 3 months, and ordered it to stop enrolling patients. As NCI wrote to the University of Pittsburgh on 29 March, "the NCI believes that the credibility of the NSABP is at stake, and the integrity of the entire NCI-supported clinical trials program may be jeopardized." NCI will decide by 1 July whether to continue funding NSABP.

–J.C.

COGs did no on-site auditing of data, a Boston University researcher, Marc J. Straus, who was participating in the Eastern Cooperative Oncology Group, was at the center of a headline-making scandal. According to FDA investigatory reports obtained through the Freedom of Information Act, Straus enrolled patients who weren't eligible for trials, failed to obtain consent from others, and falsified patient data.

Largely because of the Straus incident, NCI, in a special arrangement with FDA, agreed to start auditing COG sites at least once every 3 years. That arrangement was a compromise between "what we expected drug companies to do and what [the COGs] were doing, which was nothing," says FDA's Lisook.

A former NIH official knowledgeable about the negotiations says even that level of auditing was a bitter pill for many of the participants in the COGs. "Their reaction was, 'You guys don't trust us' and 'Why should NIH spend precious research dollars for cops?'" recalls the former official. Despite this resistance, NCI set up monitoring groups, says the official, and "we started discovering things all over the place." In particular, he recalls, the auditing found that sites could not account for experimental drugs dispensed to them.

If, in the past, NIH has reacted to each scandal in turn, creating specific solutions to solve a particular problem but not coming up with an overall response, this time the reaction may be different. As a first step, the agency is now working overtime to find out how its institutes monitor their clinical trials. Wendy Baldwin, NIH's deputy director for extramural research, has just begun to conduct a survey of monitoring practices.

And that survey could lead to action. At Varmus's behest, Baldwin has also organized a new NIH Planning Group on Clinical Trial Monitoring. This group of two dozen extramural and intramural researchers plans to hold its first meeting on 8 June, Baldwin says, to gather information about how NIH-sponsored Phase III trials are monitored and to come up with options to improve oversight. Varmus is also considering organizing a blue-ribbon panel to assess the work of this planning group and others and to make decisions about what, if anything, needs fixing.

But as more public attention is focused on clinical trials, the NIH could lose the initiative to Congress. Representative John Dingell (D–MI) has already held one high-profile hearing on the Poisson case, in which NCI received a thrashing. Dingell has scheduled another hearing for 15 June. In that hearing, and in the research community, the costs and benefits of beefed-up monitoring for clinical trials are sure to be discussed with new fervor. And at some point in the not-too-distant future, the NIH will have to decide whether the costs—in time, money, and researchers' independence—still outweigh the advantages of an overall policy, one of which might be renewed public trust in the nation's clinical trials.

–Jon Cohen