Stanford Inquiry Casts Doubt on 11 Papers

Inappropriate use of controls found in some studies of spinal fluid from psychiatric patients, but some researchers fault the investigation and question whether data were really compromised

STANFORD UNIVERSITY announced last week that 11 papers published over the past 9 years by researchers associated with the university's Mental Health Clinical Research Center may have to be clarified, corrected, or withdrawn.

Ten of the papers were tainted, the Stanford announcement said, because data from 14 patients identified in a separate study as suffering from memory loss associated with mild senile dementia were mixed with normal controls in these studies. In essence, the same patients were described in different ways in different studies. Some of these papers also contained other, unrelated, errors, Stanford said, and an additional paper contained a mistake in the way some data were reported.

The apparent contamination of the data in these papers was discovered last September during an audit of the center's work that was requested by the National Institute of Mental Health (NIMH). The matter has since been investigated by the medical school's Committee on Ethical Scientific Performance and by university provost James Rosse. NIMH is also now conducting its own investigation.

Rosse, who announced the findings last week, said his investigation found "no evidence of scientific fraud." But he termed the use of data from subjects enrolled in a study of memory loss as normal controls in other studies "a serious departure from acceptable scientific procedure."

The incident raises difficult issues concerning the rights and responsibilities of researchers who take part in collaborative projects involving several different laboratories. The 11 papers were authored by a total of 13 researchers, many of whom had little contact with, and no supervisory responsibilities for, the clinical aspects of the studies, which is where the mixture of data from experimental and control subjects apparently occurred. (Most of the authors were involved in only one or two of the studies.)

Stanford's statement said the errors might have been avoided with better interaction among the collaborators and closer supervision by senior investigators. But Jack Barchas, a coauthor on several of the papers, says that remark "is an effort to place blame across a wide spectrum of faculty members," and "has tarred a broad group of able faculty."

Just how seriously the papers were compromised by the use of control data from subjects identified as suffering from memory loss is a matter of dispute among some of the researchers involved. Stanford conducted a worst-case analysis in which the suspect data were discarded and the results of the studies were recalculated without them. On that basis, the conclusions of three papers cannot be supported, and minor corrections or clarifications are required for eight others. But one key researcher contends that the original diagnosis of the patients was incorrect; they were not suffering from senile dementia, he believes, and thus were acceptable as normal controls in other studies.

James Rosse: *Classification of the same subjects in different ways is "a serious departure from acceptable scientific procedure."*

The papers that are now under suspicion all involve an examination of the metabolites of neurotransmitters in the spinal fluids of patients suffering from a variety of mental disorders. The most prominent scientist involved in the studies is Philip Berger, a prolific researcher into the biochemical basis of psychiatric diseases, who until last year was the Mental Health Clinical Research Center's director.

Berger resigned from Stanford in May 1987 during an investigation of his management of grant money and he is now in private practice (*Science*, 31 July 1987, p. 479). The investigation resulted in the return to NIMH of $128,000 of a $5-million grant. NIMH subsequently asked Stanford to conduct an audit of the unit's annual reports over the past 9 years. It was during this audit, which was conducted by the center's current director, Adolph Pfefferbaum, that the apparent problem with the controls came to light. Pfefferbaum found a discrepancy in the number of control subjects in 3 out of 76 research protocols.

The discrepancy resulted from the inclusion in a group of controls of data from 14 elderly patients who participated in a study headed by Jerome Yesavage, a Stanford psychiatrist. The study, begun in 1978, was a preliminary investigation of the effect of a drug called nafronyl on neurotransmitter metabolites in the spinal fluid of patients with mild senile dementia. Yesavage's group made the diagnosis that the individuals were suffering mild memory loss associated with senile dementia. The study was supported in part by the drug's manufacturer, Lipha Pharmaceuticals. Berger collaborated in the study and his group performed the lumbar punctures.

Data from these subjects were stored in an extensive data bank maintained at the clinical research center, which is used to provide data for a variety of studies. At some point in the preparation of the papers now under suspicion, a decision was made to include data from these 14 subjects among data from normal controls.

The statement released by Stanford said that "like the [ethics] committee, the Provost was unable to determine exactly how the admixture of experimental patients and control subjects occurred." Berger, however, said in an interview with *Science* that he has always regarded the subjects of the nafronyl study as normal, and he therefore believes their use as controls in other studies was appropriate. Berger was a coauthor on 9 of the 11 papers.

According to Berger, the patients were reevaluated in his unit when they were referred to him by Yesavage for their lumbar punctures, and "we were unable to find problems that were not age-related." The subjects, he said, "appeared normal." Kenneth Davis, professor of psychiatry at Mount Sinai School of Medicine, who was in Berger's unit when some of the subjects

were evaluated, agrees that "they were regarded by the staff as essentially normal controls."

Nevertheless, in 1982, the same year that the nafronyl study, classifying the patients as suffering from mild dementia, was published, Yesavage and Berger coauthored another paper in which some of the same subjects were included in a normal control group. "Two investigators cannot describe them two ways in the same year," says Robert Cutler, senior associate dean of the medical school, who chaired the ethics committee.

Yesavage is currently on sabbatical in Brazil and could not be reached for comment. Cutler says, however, that Yesavage told the committee that he was not aware at the time that the subjects were included in the control group for the second study. A colleague, who calls Yesavage "a very careful researcher," says "I am convinced he was not aware that these were his own patients coming back as controls" in the second study.

Cutler said Yesavage surmised that the mixture of the data may have occurred when a research assistant put together the control group from the center's data bank. "It is possible that the nafronyl patients were there in the database as normal controls," says Cutler. One of the researchers involved in some of the suspect papers says the 14 subjects were—and still are—identified in the data bank as being different from other controls, however.

As for Berger's coauthorship of the nafronyl paper, he says that, in retrospect, the title of the paper, which included the term senile dementia, "should have bothered me more." The diagnosis reported in the paper is of mild senile dementia at the very low end of a measure that Berger now says "has never been validated."

In any case, Berger says "I didn't follow the study closely and had little interest in it." His primary interest, he said, was in collaborating with Yesavage to obtain access to normal elderly patients coming through Yesavage's unit for use as controls in a variety of other studies.

Asked to comment on Berger's explanation, Cutler acknowledged that "there is a valid question as to whether they actually had senile dementia," but he says the committee examined the records of Yesavage's initial psychometric evaluation and "the description in the [nafronyl] paper appears to be accurate."

The explanation for why Yesavage and Berger authored papers describing these subjects in different ways could therefore lie in a conflict over the diagnoses, compounded by insufficient attention to the details in each paper. Berger's judgment that they were, in fact, normal may then have prompted their inclusion in control groups for other studies.

One Stanford researcher who asked not to be identified, believes this is "exactly what happened." But, the researcher said, none of the coauthors of the affected papers who are currently at Stanford was aware of the fact that the subjects of the nafronyl study were included as normal controls until the investigation revealed the fact. Berger says "I don't specifically remember" informing other authors, but, he says, "it was common knowledge to other people on the unit that we did not consider these people to be anything other than normal controls."

In any case, Stanford's statement points out that "the fact that they had been recruited as experimental subjects would in any event make it inappropriate to include them as a group of normal controls without noting that fact." Rosse agrees. "Any publication that used these data in the way they were used should have reported that they were recruited for a different subject," he says.

Philip Berger: *"We did not consider these people to be anything other than normal controls."*

Berger concedes that "were it to be done over again, I think it would be appropriate in the description of normal controls to say that the group . . . includes patients recruited by Yesavage for his study."

The ethics committee completed its investigation in December and turned over its report to both Rosse and NIMH. In late March and early April, letters to the editors of the journals that published the 11 papers were prepared informing them that the control groups were inaccurately characterized. The letters were signed by all the relevant coauthors, including Berger.

According to Cutler, eight of the papers require relatively minor clarifications or corrections. In one other paper, the conclusions could not be supported by the data when the suspect controls were removed, but subsequent studies have suggested that the conclusion was correct, Cutler says. The remaining two papers have more serious problems. Not only were the conclusions not supported when data from the nafronyl patients were discounted, but the studies may also have been compromised because some of the patients had taken psychotropic drugs prior to the tests. So far, none of the journals have published any statements about the papers.

The incident raises the issue of the extent to which those who coauthor scientific papers should be responsible for the accuracy of all the reported data. The issue is made more complex in collaborative projects when different laboratories are responsible for different aspects of the work. Rosse, who acknowledges that the problem "is really a tough one," says "if people are going to be joint authors, they cannot deny the responsibility for all the work" presented in a publication.

Barchas, whose lab performed the biochemical analysis of spinal fluids in several of the studies, says the "absolute liability" implied by Stanford's statement "would kill collaborative research." If every person in a collaborative project is to be responsible for errors made by other, independent, participants, an outside auditor may be required to provide an "exorcism function" by examining all the data books, Barchas suggests. In fact, Barchas says this incident has already had a chilling effect on his own research; he has recently turned down collaborations with two other labs at Stanford because of problems that conceivably could arise.

Barchas says the Stanford investigators "came to us completely out of the blue" and asked to see all our records. "My records," he says, "were total and complete and correct."

The clinical aspects of the studies, Barchas notes, were handled quite separately from the biochemical analyses, although he says there was considerable interaction between himself and the group at the clinical research center. He says he approved a collaborative effort with the center because of its outstanding national reputation. In a study of this type, he says, "it is necessary that each unit be able to trust in the integrity of the other's data. . . . I hope we don't have to get into a mode where we have to audit our colleagues."

Rosse says that "there is a crucial need for explicit guidelines for laboratory research involving large research groups and multidisciplinary teams." One step, he suggests, is

for authors of papers to specifically identify the parts for which they are responsible. Rosse has asked a Faculty Senate committee to examine the issues.

In the meantime, NIMH is conducting its own investigation of the events that led to inclusion of the experimental subjects in the control groups. The Stanford investigation focused primarily on the coauthors who are still at Stanford, while the NIMH investigators are conducting more extensive interviews with those who have since left the university. Officials involved in the NIMH investigation declined to say when it is expected to be completed.

The fact that the NIMH investigation is still going on has led some to question Stanford's publication of its findings. Davis of Mount Sinai says discussion of the incident "does everybody a disservice until NIMH issues its report." Barchas also calls the Stanford statement "premature." Rosse, however, defends Stanford's decision to go public. "We think we have to manage our own affairs, and we will. We have a real concern that research coming out of this institution is credible," he says.

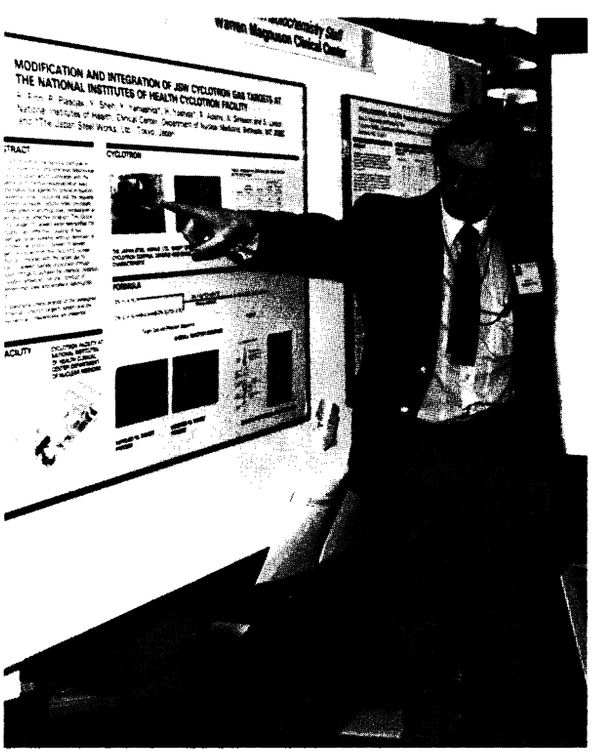
■ COLIN NORMAN

NIH Holds a Science Fair

THE FEDERAL GOVERNMENT put on a science fair last week. As with science fairs everywhere, proud scientists stood and explained their entries to judges. But these judges were not school principals and community leaders. Instead, they were executives and scientists with biomedical companies who had more than blue ribbons to offer for good projects: they held out the prospect of profitable collaborative research agreements.

The occasion was the first Industry Collaboration Forum cosponsored by the National Institutes of Health and the Alcohol, Drug Abuse, and Mental Health Administration (ADAMHA). It was the Public Health Service's first major effort to implement the Federal Technology Transfer Act of 1986. The Act allows federal labs to enter into Cooperative Research and Development Agreements (CRADAs) with industry to hasten the process of getting discoveries out of the labs and into the marketplace.

Under the Act, federal labs may provide personnel, services, and property—but not funding—toward a joint research project with a commercial firm. The firm can contribute cash and other services and materials, but cash alone is not enough. The firm has to make an "essential . . . or very important" contribution to the project, said Philip S. Chen, Jr., associate director for intramural affairs at NIH. "We do not wish to use the Cooperative Research and Development Agreement mechanism as simply a cheap way to buy access into the federal laboratory research enterprise," Chen said.



Ernie Branson, NIH

Show and tell. *Ronald D. Finn of the NIH Clinical Center displays his poster on cyclotron research for industry observers.*

For its financial and other investments, the outside firm receives an exclusive license to any marketable invention that results from the project, although the government retains the right to use the invention without paying a licensing fee. A major advantage to the commercial firm is that it need not go through the arduous process that the Commerce Department uses for licensing government patents, a process so unwieldy and slow that few government patents have been licensed to industry.

The government also benefits. Its scientists get to keep a share of royalty income and the lab sponsoring the research gets to keep royalties that once would have had to be returned to the federal treasury.

The forum extends the growing movement at NIH to foster closer ties with industry. That movement began 3 years ago when NIH director James B. Wyngaarden issued liberalized work rules to allow NIH scientists to consult with commercial companies on their own time, and for pay. That rule was promulgated largely to help government scientists supplement salaries that were not keeping up with those offered in the private sector. But it also had the effect of spawning some joint research programs.

Some 38 collaborative projects already exist. They include Robert Gallo's research on AIDS vaccine development with IMMUNO Co., a joint project with Sandoz on the use of cyclosporin in multiple sclerosis, and one with Hybritech on the use of monoclonal antibodies in the treatment of lung cancer. The forum was designed to foster more collaborations, and NIH encouraged scientists to participate.

"We were encouraged all right," one scientist present groused. "They even twisted the arms of some of us."

More than 200 industry representatives showed up to peruse the 105 posters on topics ranging from acne to vaccines for pertussis. Industry participants paid $150 each to attend, mostly to cover the cost of materials and the use of the hotel where the conference was held. Chen said NIH probably lost money on the deal, since 400 registrants were needed to break even.

But Chen remained pleased with the turnout, and several poster-guarding scientists expressed amazement that so many companies were interested in seeing the fruits of the nation's largest biomedical research effort. Chen said future forums would be smaller, focused on one area of research, and held at the NIH campus to reduce the costs.

Chen admits that several major issues need to be worked out, including potential conflicts of interest for government scientists. Draft rules for government scientists participating in joint ventures are being discussed now, he said. Another is the commercialization of inventions discovered under joint research. NIH and ADAMHA, as publicly funded basic research institutions, have rarely had to deal with the thorny questions of whether products resulting from their research are priced affordably and made available to those groups who need them most. Also, who is liable for injuries or loss resulting from use of a product developed in a joint venture? The answers to these questions may well determine the success of the NIH/ADAMHA venture into the marketplace.

■ GREGORY BYRNE