Controversy Erupts Over Proteomics Studies

BY MICHELE G. SULLIVAN
Mid-Atlantic Bureau

Systematic bias in the design of several underlying studies raises doubt over whether a serum proteomics test based on those studies can accurately identify ovarian cancer, two independent biostatisticians have argued.

The researchers, both of the University of Texas M.D. Anderson Cancer Research Center, Houston, have been unable to reproduce the high sensitivity and specificity rates reported in a 2003 study of the technique (J. Natl. Cancer Inst. 2005;97:307-9).

The problem, said Keith A. Baggerly, Ph.D., and Kevin R. Coombes, Ph.D., lies not in the fundamental concept—that cancer-shed proteins in serum may be able to identify patients who have even very early-stage cancer—but in the way the data sets were processed in both the 2003 study and the original 2002 National Cancer Institute (NCI) study upon which it was based.

"We're not saying proteomics doesn't work," Dr. Baggerly said in an interview. "It may very well work. But these data sets can't be used to say this approach works."

The method involves using mass spectroscopy to display proteins in serum as a series of peaks and valleys of varying strength. A computer-driven mathematical algorithm finds unique patterns expressed in the serum of patients with the disease. Several researchers are investigating proteomics' application in ovarian cancer, using different algorithms and spectrometers. All of the decoding work is being performed on three publicly available sets of spectral data, which were processed as part of the original proof-of-concept study by NCI researchers led by Emmanuel F. Petricoin III, M.D. (Lancet 2002;359:572-7).

Dr. Baggerly and Dr. Coombes reanalyzed the data used in a 2003 paper by Wei Zhu, Ph.D., and associates, of the State University of New York at Stony Brook. By using the same NCI data sets—samples from women with ovarian cancer, women with benign ovarian cysts, and healthy controls—but a new protein-recognition pattern, Dr. Zhu achieved perfect discrimination (100% sensitivity, 100% specificity) of patients with ovarian cancer, including early-stage disease, from normal controls (PNAS 2003;100:14666-71). Dr. Zhu's results were even better than those originally reported by Dr. Petricoin and colleagues in their 2002 study.

When Dr. Baggerly reanalyzed the Zhu data, he was unable to arrive at the same results. The Zhu study identified a pattern involving 18 protein peaks that separated controls from cancers. For Dr. Baggerly, the pattern resulted in significant accuracy in the first data set, which contained serum from all three groups, but not in the second data set, which contained only serum from cancer patients and healthy controls.

In the second data set, 13 of the 18 peak differences changed signs—that is, peaks associated with cancer in the first group were associated with controls in the second group, and peaks first associated with controls switched to cancers. "This reversal isn't consistent with a persistent difference between cancer samples and control samples," Dr. Baggerly said.

The researchers then chose 18 random protein peaks from the same regions of spectral data as Dr. Zhu's peaks. The random peaks separated cancer samples from controls up to 56% of the time, depending on the strength of the signals used. Because the pattern of protein expression was inconsistent between the data sets, they concluded, the values did not represent biologically important changes in the serum of cancer patients.

The problem, Dr. Baggerly asserts, is that Dr. Zhu processed the serum samples in a nonrandomized way that the spectra were acquired in the initial study by Dr. Petricoin and his colleagues.

"They ran all the controls on one day and all the cancers on the next day," Dr. Baggerly said. "This is the worst kind of design when you are using a machine that can be subject to external factors," such as changes in calibration or mechanical breakdown.

In fact, he said, a June 2004 study in which Dr. Petricoin participated also suffered from such a problem (Endoc. Relat. Cancer 2004;11:163-78). This study used a different mass spectrometer, which began to break down on day 3 of running the samples.

In a letter to the editor, Dr. Petricoin admitted the problem, but said, "We cannot detect whether the cancer data acquired on the previous day were convincingly negatively affected by the spectrometer failure."

Dr. Baggerly contends that a better design involving randomizing sample processing would allow separation of differences due to biology from those due to external factors.

His failure to find reproducibility does not surprise Dr. Petricoin and his colleague, Lance A. Liotta, M.D., who participated in the 2002 and 2004 studies. Their commentary appears in the same journal. Each of the data sets, all of which are available without restriction online, was generated with different machines and methods to test those machines and methods.

"We would be surprised if the experimentally designed process changes between these two studies did not result in altered spectra. In fact, a goal of these experiments was to study the spectral alterations produced by changing the process," they said.

Because serum proteomics is in its infancy, they wrote, there is no procedure to standardize intra- and inter-laboratory comparisons. Only after that standardization happens can well-designed, meaningful, and reproducible studies be conducted.

In the meantime, they concluded, researchers who wish to attempt such studies should keep open lines of communication with those who originally produced the data. "A meaningful analysis of reproducibility requires communication. … Without such communication, data can be misinterpreted; unwarranted, overextended conclusions can be drawn; and misinformation can be spread," he said.

Additional research using meticulously designed studies is needed, Dr. Baggerly said. "If a test for ovarian cancer ever does come about from these data, I'd need to see a lot more studies before I'd send my mother out to get it." ■

OvaCheck Unaffected, Developer Says

The serum proteomics study design debate won't affect the progress of OvaCheck, a proteomics test being developed as a screen for women at high risk for ovarian cancer, said Peter Levine, head of the Maryland firm developing the test.

"This is a purely academic debate," said Mr. Levine, chief executive officer of Correlogic Systems Inc. "It has no bearing whatsoever on the state of the development of the technology today or on any of the other work researchers have been pursuing in this field."

OvaCheck uses a sophisticated mathematical algorithm and mass spectroscopy to identify a specific pattern of serum proteins associated with even very early-stage ovarian cancers. The method was based on a 2002 National Cancer Institute (NCI) study, but it uses a different mass spectrometer and different spectral signals to identify cancer samples. Correlogic Systems is conducting validity testing on hundreds of samples but has not released any data on those tests.

The study design debate adds nothing to the development of proteomics technology because it focuses on outdated research, Mr. Levine said. "These studies are 2 and 3 years old," he said. "Since then, scores of additional papers have been published on this technique and various other techniques."

In fact, reanalyzing older studies may put forth the mistaken impression that serum proteomics has no future as a screening or diagnostic tool.

"The [NCI study] ushered in a revolution in the way we look at this biological data," he said. "But it was just a proof-of-concept study. No one ever claimed it was a test for ovarian cancer."

Many additional, more recent studies continue to expand on this original idea, including the research Correlogic Systems is performing, Mr. Levine said.

"We are refining our own technology as we go through the testing process, and that kind of research and development—tweaking the equipment and the process—goes on forever, as it should. Continuing to debate these early papers is like doing a thesis on the Wright brothers' first flight, when you already have a 747 that flies."

OvaCheck, however, is still struggling through administrative processes at the Food and Drug Administration. Correlogic Systems hoped to license OvaCheck as a lab-developed test regulated under the Clinical Laboratory Improvement Amendments (CLIA). But the FDA determined last year that the software powering OvaCheck is a medical device covered by interstate commerce regulations and thus subject to FDA premarket review.

"We're still working with FDA on that issue and hope to have it resolved soon," Mr Levine said.

Chronic Visceral Pelvic Pain Responds to Neurostimulation in Small Study

PALM SPRINGS, CALIF. — Pain scores declined by half and opioid use significantly declined in a small number of patients treated with spinal cord stimulation for chronic visceral pelvic pain at the Cleveland Clinic Foundation.

Nagy Mekhail, M.D., chairman of pain management, and his associate Leonardo Kapural, M.D., conducted a small clinical study of spinal cord stimulation for pelvic pain based on recent studies implicating dorsal column pathways in the transmission of visceral pelvic pain.

Six patients were enrolled in the study Dr. Mekhail presented at the annual meeting of the American Academy of Pain Medicine.

All had long-standing histories of chronic pelvic pain of a mean duration of nearly 15 years. All had pelvic adhesions and had undergone multiple surgical explorations for endometriosis and other diagnoses. Their clinical diagnoses included chronic pelvic inflammatory disease, adenomyosis, urinary retention, dyspareunia, vaginal prolapse and vulvodynia, and vulvar vestibulitis. Multiple surgical and medical therapies, including antidepressants and opioids, failed to relieve their pain for long.

After psychological evaluation and approval by an implantable devices committee, each patient underwent a spinal cord stimulation trial of 7-14 days and was implanted with two dual compact or dual quad leads connected to a generator. The lead tip was at the T11 or T12 level.

After an average follow-up of more than 30 months, the patients' mean visual analog score declined from 8 to 3, with all patients reporting a 50% or greater decrease in pain. Mean Pain Disability Index scores declined from 57.7 to 19.5, and mean opioid use was reduced from 22.5 mg morphine equivalent per day to 6.6 mg per day.

Dr. Mekhail acknowledged that the study was very small and the results were preliminary. However, he said these encouraging results suggest that spinal cord stimulation may offer "significant therapeutic potential" for difficult to treat patients with severe, long-standing visceral pelvic pain.

—Betsy Bates