COMPUTERS:
Scientific friend or foe?

Whether by mistake or fraud, the misuse of technology can sabotage important research

By Dan Vergano
USA TODAY

It's an old joke among techie types: To err is human. To really screw up, you need a computer.

Now scientists are learning that lesson, too. But they don't think it's funny.

As computers extend their reach into the research laboratory, they are making possible new kinds of science, but they're also creating pitfalls along the way. Some computer programs are causing problems on their own, wreaking unexpected havoc with scientific data. In other cases, scientists are using computers -- either on purpose or out of ignorance -- to deceive.

In June, a *BMC Bioinformatics* journal report detailed how the widely used spreadsheet Excel can turn gene names into dates, ruining the ability to call up or analyze data correctly. One simple example: A default date converter altered the name DEC1 — short for the tumor-suppressor gene Deleted in Esophageal Cancer 1 — into "1-DEC." Alarmed gene researchers rocketed the report to the top of the "most-viewed" list for BioMed Central, publisher of about 150 "open-access" journals. Some quickly noted that the same formatting errors occur in other spreadsheets as well, corrupting the names of many genes used in experiments.

"As you can imagine, it's a problem when we have particular genes to find for basic medical purposes," says report senior author John Weinstein of the National Cancer Institute. His co-authors who first spotted the error, Barry Zeeberg of the NCI and Joseph Riss of the Laboratory of Biosystems and Cancer, went through more than a little frustration figuring out the problem. "I shudder to think how many pieces of work this might have affected," wrote another researcher in comments discussing the paper online.

Dependent on technology

Computers have been scientific workhorses since the Army's ENIAC prototype computer made calculations for the Manhattan Project in World War II. But today, reliance on computers has turned into dependence on programs like Excel, Weinstein says. Where once gene researchers only looked at a few genes in experiments, they now may look at tens of thousands, something only possible with sophisticated spreadsheet and database programs.

And scientific institutions such as NASA, which teamed up last month with Silicon Valley firms to create a "Space Exploration Simulator" supercomputer, crave ever-more-powerful computers to solve important research questions.

But a few recent speed bumps illustrate computer concerns for scientists:

▶ In July, the editors of the *Journal of Cell Biology* cautioned researchers against misuse of the popular photo-manipulation program Photoshop to clean up images, a practice that can cross the line into misconduct. For example, some researchers were removing imperfections in photos of gels used to find genes. "Just because the tools exist to clean up sloppy work digitally, that is no excuse to do sloppy work," *JCB* editors warned.
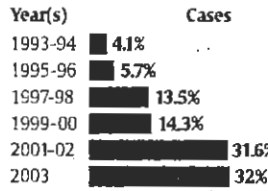
▶ In January, a computer glitch nearly disabled NASA's Mars rover Spirit, leaving it incommunicado with mission controllers for two weeks. Rewritten computer codes updated the flash memory on the mobile probe and put it back into service.

▶ Last year's report on the space shuttle Challenger accident that killed seven astronauts excoriated "Engineering by Viewgraphs." Report authors concluded that NASA scientists had used PowerPoint presentations to make it look as if they had done analyses that they hadn't. Citing a criticism made by Yale's Edward Tufte, an information presentation expert, the report authors wrote that over-reliance on presentations instead of analysis was seen as "an illustration of the problematic methods of technical communication at NASA" that contributed to the shuttle disaster.

"What we're seeing here are surface issues that point to deeper

> "What we're seeing here are surface issues that point to deeper problems with the way we use existing software to present and handle data."
>
> — Leo Herbette
> of Exploria Inc.

problems with the way we use existing software to present and handle data," says Leo Herbette of Exploria Inc., a Hartford, Conn., firm that develops software to optimize the presentation of scientific and medical information.

With respect to researchers "Photoshopping" their data, for example, "the vast majority of the (photo) manipulations that we see are due to a lack of knowledge about what is appropriate," says *JCB* editor Mike Rossner. Scientists have always processed images to some extent, he says, but standards for digital photos are still evolving.

Is it an age thing?

Senior researchers tend to be less adept with computers than younger ones, which may add to problems in developing standards, say John Krueger of the Public Health Services' Office of Research Integrity. Old-fashioned mentoring about what is acceptable behavior in handling and in presenting may go missing because of the generation gap in computer ability. The mentor no longer remains as engaged, from start to finish, in the presentation of data.

Such a scenario played out in the 2002 case of physicist Hendrik Schoen of Bell Labs. According to a lab report, he fabricated data and images in 17 papers published on molecular electronics, despite the nominal oversight of his mentor.

ORI has seen an increase in photo manipulation cases, from 14.3% of scientific misconduct cases in 1999-2000 to 32% of cases in 2003 alone. However, increased computer power also has made catching instances of research fraud easier, says Krueger, so the impact tends to even out. "We're also seeing increased computing power make institutions better at detecting falsified images after suspicion arises about a possible (misconduct) problem."

When investigators take a hard look at the data, faked pictures are easily spotted, he says.

Handcuffed by computers

Some critics say that current computing tools may be holding back thinking on big problems. Programs that present data one slide or one photo at a time, Herbette says, fail researchers by not letting them make links between the data generated by disparate experiments, the real promise of the computer age.

Weinstein says the issue is that medical researchers are in transition from an era of single-experiment-driven research to a computer-driven one that generates new data, ideas and methods to investigate biological phenomena.

Obviously, programs that cause errors in databases are a disaster, he says, but more important, researchers need to recognize that computers aren't just tools but increasingly are intrinsic to the science itself.

The growth of "omic" sciences — disciplines like genomics, proteomics and others -- rests entirely on the ability of computers to draw connections between bits of data in a huge sea of results generated by experiments.

The National Institutes of Health inaugurated a 10-year "NIH Roadmap" program last year to deal, in part, with the demands of these new disciplines. But journal editors who publish scientific studies and grant institutions that fund them very often see less value in efforts to build and analyze scientific databases than in old-fashioned experiments.

That has to change, Weinstein says, to make the promise of the human genome and related science a reality.

Some predict that by 2010 or 2015, most practicing biomedical researchers will not be at the lab bench at all — they will instead be sitting at their computers mining databases generated by biological factories.

"I think that's mistaken," Weinstein says. "But it's clear that information science is becoming a larger and larger part of biological research, and we have to make room for it."