NEWS

The New Math of Clinical Trials

Other fields have adopted statistical methods that integrate previous experience, but the stakes ratchet up when it comes to medical research

HOUSTON, TEXAS—If statistics were a religion, Donald Berry would be among its most dogged proselytizers. Head of biostatistics at the M. D. Anderson Cancer Center here, he's dropped all hobbies except reading bridge columns in the newspaper. He sends out e-mail missives at 3:00 in the morning. The running joke in the department is that Berry, his curly gray hair perpetually tousled, never sleeps. Admittedly, sleep doesn't come easily to a man on a mission.

Berry, 63, adheres to a branch of statistics named after an 18th century minister, Thomas Bayes, whose followers advocate incorporating prior knowledge into experiments and sometimes altering them as they run to take into account accumulating results. Although Bayesian designs are now widely used in everything from astrophysics to ecology (*Science*, 19 November 1999, p. 1460), they've been slower to catch on in medical research, particularly clinical trials. That's where Berry comes in.

A Bayesian since the 1960s, Berry for years was unable to implement his unorthodox approach. Then, in 1999, he was offered a golden opportunity: Come to M. D. Anderson, one of the largest cancer centers in the United States, with a reputation for being the "Wild West" of oncology research, and transform how it designs and runs many of the 800 clinical trials being conducted at any given time.

Berry's perch at Anderson has fueled his resolve to spread the Bayesian word. He crisscrosses the country speaking with cancer advocates, drug companies, and the Food and Drug Administration (FDA); the latter is beginning to consider Bayesian trials in new drug applications and is planning a May meeting on the subject.

His critics, however, hope his ideas won't take hold. Berry's skepticism that mammograms help younger women left him accused of risking lives; his approach to clinical trials has prompted worries about bad drugs slipping through the system. Bayesian drug studies risk "saying [a treatment is] positive too often," says biostatistician Stephanie Green of the Fred Hutchinson Cancer Research Center in Seattle, Washington. But critics and supporters alike have a grudging admiration for Berry's persistence. "He isn't swayed by the status quo, by people in power in his field," says Fran Visco, head of the National Breast Cancer Coalition in Washington, D.C. "You have to respect him for that," she adds, "whether you agree with him or not."

Maverick beginnings

Berry stumbled into statistics after an erratic college career. He skipped classes regularly and took a 3-year break, in 1960, to volunteer for the army. By his senior year, he and his wife had four sons (two more children, both girls, would follow), and Berry had little idea what to do with his life. A professor suggested statistics; Berry took the advice and enrolled in graduate school at Yale University. After completing his dissertation in 1971, he moved to the University of Minnesota.

From the start, Berry was drawn to the Bayesian school of thought, then widely viewed as an oddity within the field. The Bayesian approach calls for incorporating "priors"—knowledge gained from previous work—into a new experiment. "The Bayesian notion is one of synthesis … [and] learning as you go," says Berry. He found these qualities immensely appealing, in part because they reflect real-life behavior, including the way doctors practice medicine.

Bucking tradition. Donald Berry's support for Bayesian designs is changing the face of clinical trials, especially at his home base of M. D. Anderson Cancer Center.

CREDIT: BARRY G. SMITH

But learning as you go collides with the decades-old clinical trials paradigm. To guard against bias—from doctors, drug companies, and even patients—each phase of a traditional clinical trial is run from start to finish without interference from interested parties. Outside scientists monitor the data regularly; although a trial can be stopped early if patients appear unduly harmed or helped by the new treatment, the protocol itself can't normally be changed.

A Bayesian approach demands more than sporadic monitoring-board meetings, however. Bayesian trials often unveil data while a study is ongoing. What's more, researchers can use these early results to reallocate patients to different treatment groups depending on how the first batch of patients, or even a single patient, fares. Berry also favors other approaches foreign to clinical trials, such as answering questions about multiple drugs in a single experiment, a method known as factorial design. Factorial designs include a treatment arm for every drug combination possible, leading to unwieldy experiments whose results can be tough to interpret.

Some doctors agree with Berry that the standard approach to clinical trials is problematic. Elihu Estey, who oversees the treatment of acute leukemias at Anderson, points out that the typical paradigm assigns patients to different study arms with equal probability, even in the face of mounting evidence that one arm offers a better shot at survival. "The patients themselves, if they knew the way the trials are conducted, wouldn't be too thrilled," he says.

A big break for Berry came in 1990, when he was invited to join Cancer and Leukemia Group B (CALGB). It's one of the country's 10 cooperative groups: multiinstitutional collaborations on large-scale cancer clinical trials. Berry would be the lead statistician for CALGB's breast cancer studies. He was not greeted warmly.

"I objected rather strenuously," recalls I. Craig Henderson, a breast oncologist at the University of California, San Francisco, who had heard that Bayesians were "loosey-

goosey" in adhering to the rules.

Henderson subsequently had a change of heart: Last year, he was the first in a string of authors on one of the largest breast cancer studies Berry has designed, with more than 3000 women. Its factorial design revealed that adding the drug paclitaxel (Taxol) to standard chemotherapy is beneficial, and that high doses of doxorubicin (Adriamycin), one of the most toxic chemotherapy agents, don't fight cancer any more effectively than lower doses. This came as a great surprise, and some criticized the study for its unusual methodology.

Despite Berry's relentless efforts to convert Bayesian nonbelievers in CALGB, the group has yet to conduct a fully Bayesian study. But CALGB and other cooperative groups are adopting factorial designs to answer more questions, more quickly. "Maybe some of these designs don't give you the absolute perfect answer," says Eric Winer, co-chair of CALGB's breast committee and an oncologist at Dana-Farber Cancer Institute in Boston. But, he adds, "it may be good enough if the alternative is waiting another 10 years."

A nose for controversy

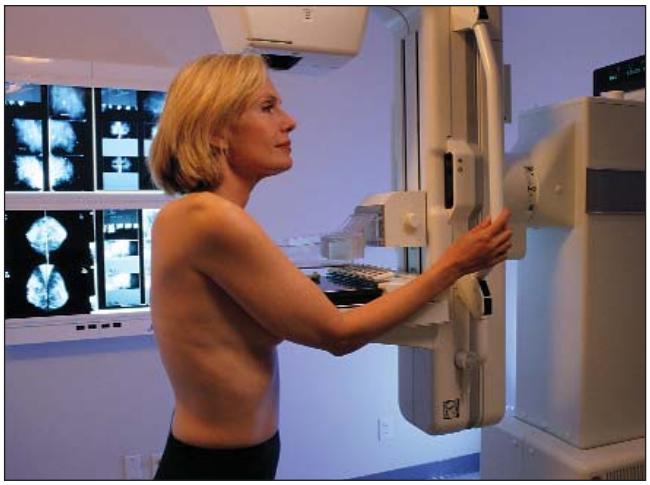
Over dinner at an Italian restaurant in Houston, Scott Berry, 37, munches on spinach pizza and considers why his father leaps into one controversy after another. There's one thing, he says, that he's certain of: "He doesn't do it for the notoriety."

Notoriety, however, is something Don Berry has amassed in impressive quantities over the years. In the late 1990s, he became a lightning rod in the debate over whether women under 50 benefit from mammograms. As co-chair of a National Institutes of Health panel on the subject, he reported that regular screenings of 2500 women under 50 would be needed to extend the life of one. "I focused very much on the risks" of mammography, Berry says, including false positives and finding tiny tumors that are unlikely to spread.

In 2002, Berry testified in Congress on the subject; now–Senate Majority Leader Bill Frist (R–TN) minimized Berry's findings because he lacked an M.D. degree. Berry received death threats, including one from a man who believed his wife's life had been saved by a mammogram. He was also accused of rampant sexism. "People said, 'It's because you're a man; if this were prostate cancer, it would be different,' " he recalls. What his critics didn't know, he says, is that he feels even more strongly about routine testing for prostate-specific antigen (PSA), a controversial marker of prostate cancer. Berry doesn't know his PSA level and has no interest in learning it; like mammography, he believes, its inappropriate use leads to more heartache and unnecessary invasive procedures than benefits.

Fervent belief that he can shift medical opinion has sustained Berry through some dispiriting times. It's the same kind of determination that's kept him bicycling religiously each day from his home to his office, even during Minnesota winters when temperatures plunged below –30°C and hot breath turned his beard to ice and glued it to his ski mask.

One of his lowest points came in the mid-



When to start? Berry and a cadre of others argue that, for women under 50, mammograms bring more harm than gain.

1980s. Researchers at the University of Michigan were testing a new technology called extracorporeal membrane oxygenation (ECMO) on desperately ill infants. ECMO takes over the job of oxygenating the blood and gives the struggling heart and lungs time to grow or heal. The Michigan study reallocated infants to standard therapy or ECMO depending on how previous babies in that same study had fared. The result was a sharply skewed trial: 11 babies in the ECMO treatment group, all of whom survived, and one baby in the standard arm, who died.

Doctors and statisticians roundly criticized the trial, arguing that one baby in a control group wasn't sufficient to come to any conclusion. Although Berry did not participate in the trial, he conducted his own analysis. That prompted him to defend the research, and he countered that 100% survival in the ECMO group was remarkable given the high death rates observed in similar babies in the past.

Berry had little influence, however, and a team at Harvard launched a more standard ECMO trial. Berry publicly accused the Harvard researchers of killing babies, a belief he maintains to this day. That trial found ECMO to be vastly superior, and today the machines are regularly used on infants and children. Berry, however, was left deeply disheartened by his inability to prevent the Harvard study, as well as a subsequent one in the United Kingdom. In all, 58 infants—57% of those allocated to standard care—died; of those allocated to ECMO, 31 infants, or 25%, died.

Outside M. D. Anderson, true Bayesian clinical trials remain rare. "We don't use Bayesian designs here because I think the system works reasonably well without them," says David Harrington, the lead biostatistician at Dana-Farber. Both Harrington and Ross Prentice, a biostatistician at the Fred Hutchinson Cancer Research Center, say that a well-designed Bayesian trial should reach the same conclusion as traditional methods. But, says Prentice, he worries that when it comes to Bayes, "are you making more assumptions, [and] are those assumptions having more weight than they should?"

The integration of priors into a Bayesian design is among the most deep-seated concerns cited. Some worry that priors could perpetuate incorrect or anomalous early data. David Spiegelhalter, a Bayesian statistician at the Medical Research Council in the United Kingdom, admits that he's seen some disastrous Bayesian analyses that do just that. One example he cites is a quality-control comparison that may have erroneously promoted one hospital over another. The drug field, he says, still has few Bayesian analyses, but "I'm dreading the first time something high-profile comes up that hasn't been done well."

Preaching the word

At Anderson, evidence of Berry's influence is plain. His department exploded from a dozen people to 133. (Many although not all of the 60-odd statisticians are Bayesian.) Anderson now insists that companies or hospitals collaborating with it on trials examine the data throughout. Berry also presses hard to spread Bayesian teachings outside Anderson's walls; his close colleague Peter Thall co-taught a 3-day course in December to 100 FDA statisticians.

Bayesian believers like Berry and Thall know that getting FDA's stamp of approval will be crucial. But although FDA regulators who approve devices have long accepted Bayesian designs, the drug staff remains uncertain.

Rumors at Anderson about FDA "closet

CREDIT: PETE SALOUTOS/CORBIS

Bayesians" notwithstanding, the agency so far has approved only one drug on a Bayesian platform. It's a pill that combines pravastatin (Pravachol), an existing cholesterol-lowering drug, with aspirin. Approval was based in part on a Bayesian analysis that made it easier to synthesize information from five previous trials, and to allow for diverse sets of patients within each of those studies. FDA approved the combination drug in June 2003.

Scott Berry tells the pravastatin story with pride. Father and son launched Berry Consultants in 2000, and it worked with the drug's manufacturer, Bristol-Myers Squibb, to shepherd it through approval. "Most of our meetings take place between 12 and 1 a.m.," says Scott, who's the company's sole full-time employee.

Because of potential conflicts of interest with Anderson, Berry Consultants rarely advises on cancer. The overwhelming majority of its business is medical, however, such as helping the device firm Medtronic gain approval for an improved shunt for infants with hydrocephalus. Some companies seek out Berry Consultants in the wild hope that a drug or device that's performed poorly in traditional trials can somehow undergo a Bayesian resurrection. (Such a "rescue analysis" is rarely a possibility, both Berrys agree.)

His colleagues may be nearing retirement, but Don Berry isn't ready to slow down anytime soon. He's been a workaholic for as long as Scott can remember. All four of the Berry boys played ice hockey as children, and Scott remembers his father attending games back in the 1970s with work and a clipboard in hand. Goalie Scott would see his father watching as the puck slid toward his net. But once it glided safely away, Scott would glance up and spot his father braced against the clipboard, scribbling away.

–JENNIFER COUZIN

NEWS

Making Sense of a Heart Gone Wild

Armed with computer models, interdisciplinary teams of researchers are studying what triggers life-threatening fibrillation—and the even deeper mystery of why it can be stopped

Richard Gray, a biomedical engineer at the University of Alabama, Birmingham, studies the heart for a living, but last year the heart's mysteries struck close to home. Gray's 68-year-old father called 911, complaining of chest pain. The paramedics were already on the scene when he suddenly collapsed. He had gone into ventricular fibrillation—his heart running amok, its muscle fibers all marching in time to their own drummers instead of beating in unison.

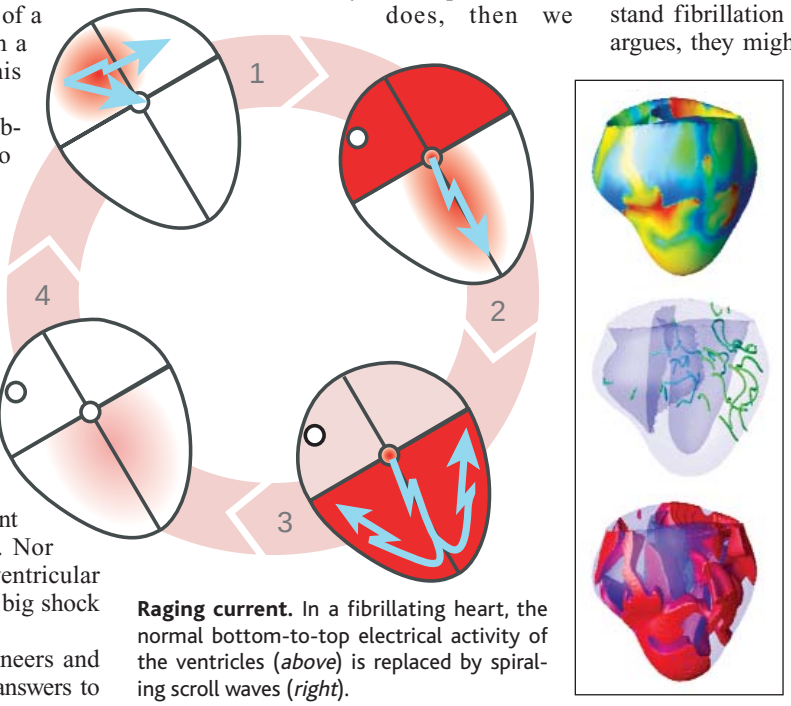
Ventricular fibrillation is a death sentence if not treated within 10 minutes, but John Gray was in luck. A member of the rescue squad applied the paddles of a defibrillator to his chest and with a whomp of electricity shocked his heart back into its normal rhythm.

Hundreds of times a day, defibrillation resuscitates people who would otherwise die in minutes. For implantable cardioverter defibrillators (ICDs), the success rate exceeds 99%. (External defibrillators, like the one used by the rescue squad, have a lower success rate, primarily because they are not always applied in time.) It's a true medical miracle—and as befits a miracle, no one can explain why it works. "We don't even know how the electric current goes into the heart," says Gray. Nor does anyone really know how ventricular fibrillation gets started, or why a big shock brings it to an end.

Gray is one of many bioengineers and heart specialists who expect the answers to emerge from mathematical models of the heart. Researchers are experimenting with virtual hearts in part because it is easier than tinkering with a living, beating one. And there is no way to look beneath the surface of a real animal heart. As Alan Garfinkel, a cardiologist at the University of California, Los Angeles, puts it, "You can't get the light into the meat."

So far, mathematics has answered some questions but raised others. James Keener, a mathematician at the University of Utah, Salt Lake City, says that if defibrillation worked the way most experts think it does, then we would have a lot more dead patients. "If we invoke the prevailing theory, the probability of success is no greater than 20% in our numerical simulations—regardless of the amplitude of the shock," says Keener. "Yet defibrillators have a success rate that approaches 100% as the shock gets larger. So we have a problem."

Some people might argue that this is a good problem to have. If the treatment works, who cares that no one understands why? Garfinkel, for one: "I would urge that electrical defibrillation, the delivery of a huge, painful shock by an implanted $40,000 device, is neither a medically satisfactory solution, nor does it represent any scientific insight into the phenomenon," he says. If cardiologists could understand fibrillation from first principles, he argues, they might be able to improve the treatment with less expensive equipment, less painful and damaging shocks, and potentially with antiarrhythmic drugs, which have until now been an embarrassing flop.



Raging current. In a fibrillating heart, the normal bottom-to-top electrical activity of the ventricles (*above*) is replaced by spiraling scroll waves (*right*).

CREDIT: (INSET) F. XIE, Z. QU, J. YANG, A. BAHER, J. WEISS, A. GARFINKEL, *JOURNAL OF CLINICAL INVESTIGATION* (IN PRESS, MARCH 2004)

The mathematical heart

Mathematical models showed long ago that there is some method to the apparent madness of the fibrillating heart. Ventricular fibrillation is first and foremost a malfunction in the heart's electric circuitry. In a normal heartbeat, electrical activity starts near