The New York Times

July 13, 2013

Do Clinical Trials Work?

By CLIFTON LEAF

EVERY spring, some 30,000 oncologists, medical researchers and marketers gather in an American city to showcase the latest advances in cancer treatment.

But at the annual meeting of the American Society of Clinical Oncology last month, much of the buzz surrounded a study that was anything but a breakthrough. To a packed and whisper-quiet room at the McCormick Place convention center in Chicago, Mark R. Gilbert, a professor of neuro-oncology at the University of Texas M. D. Anderson Cancer Center in Houston, presented the results of a clinical trial testing the drug Avastin in patients newly diagnosed with glioblastoma multiforme, an aggressive brain cancer. In two earlier, smaller studies of patients with recurrent brain cancers, tumors shrank and the disease seemed to stall for several months when patients were given the drug, an antibody that targets the blood supply of these fast-growing masses of cancer cells.

But to the surprise of many, Dr. Gilbert's study found no difference in survival between those who were given Avastin and those who were given a placebo.

Disappointing though its outcome was, the study represented a victory for science over guesswork, of hard data over hunches. As far as clinical trials went, Dr. Gilbert's study was the gold standard. The earlier studies had each been "single-arm," in the lingo of clinical trials, meaning there had been no comparison group. In Dr. Gilbert's study, more than 600 brain cancer patients were randomly assigned to two evenly balanced groups: an intervention arm (those who got Avastin along with a standard treatment) and a control arm (those who got the latter and a placebo). What's more, the study was "double-blind" — neither the patients nor the doctors knew who was in which group until after the results had been assessed.

The centerpiece of the country's drug-testing system — the randomized, controlled trial — had worked.

Except in one respect: doctors had no more clarity after the trial about how to treat brain cancer patients than they had before. Some patients did do better on the drug, and indeed, doctors and patients insist that some who take Avastin significantly beat the average. But

the trial was unable to discover these "responders" along the way, much less examine what might have accounted for the difference. (Dr. Gilbert is working to figure that out now.)

Indeed, even after some 400 completed clinical trials in various cancers, it's not clear why Avastin works (or doesn't work) in any single patient. "Despite looking at hundreds of potential predictive biomarkers, we do not currently have a way to predict who is most likely to respond to Avastin and who is not," says a spokesperson for Genentech, a division of the Swiss pharmaceutical giant Roche, which makes the drug.

That we could be this uncertain about any medicine with $6 billion in annual global sales — and after 16 years of human trials involving tens of thousands of patients — is remarkable in itself. And yet this is the norm, not the exception. We are just as confused about a host of other long-tested therapies: neuroprotective drugs for stroke, erythropoiesis-stimulating agents for anemia, the antiviral drug Tamiflu — and, as recent headlines have shown, rosiglitazone (Avandia) for diabetes, a controversy that has now embroiled a related class of molecules. Which brings us to perhaps a more fundamental question, one that few people really want to ask: do clinical trials even work? Or are the diseases of individuals so particular that testing experimental medicines in broad groups is doomed to create more frustration than knowledge?

Researchers are coming to understand just how individualized human physiology and human pathology really are. On a genetic level, the tumors in one person with pancreatic cancer almost surely won't be identical to those of any other. Even in a more widespread condition like high cholesterol, the variability between individuals can be great, meaning that any two patients may have starkly different reactions to a drug.

That's one reason that, despite the rigorous monitoring of clinical trials, 16 novel medicines were withdrawn from the market from 2000 through 2010, a figure equal to 6 percent of the total approved during the period. The pharmacogenomics of each of us — the way our genes influence our response to drugs — is unique.

HUMAN drug trials are typically divided into three phases. In the first, researchers evaluate the safety of a new experimental compound in a small number of people, determining the best way to deliver it and the optimal dosage. In Phase 2, investigators give the drug to a larger number of patients, continuing to monitor its safety as they assess whether the agent works.

"Works" in this stage is broadly defined. Seeing that the drug has any positive effect at all —

say, that it decreases the level of a blood marker associated with a disease — is often enough to move a drug to Phase 3. Even so, most experimental drugs fail before they get to Phase 3.

The few that make it to Phase 3 are then tested for safety and efficacy in hundreds or thousands of patients. This time, the outcomes for those taking the new drug are typically compared head-to-head with outcomes for those getting a placebo or the standard-of-care therapy. Generally, the Food and Drug Administration requires that two "adequate and well-controlled" trials confirm that a drug is safe and effective before it approves it for sale, though the bar can be lower in the case of medicines aimed at life-threatening conditions.

Rigorous statistical tests are done to make sure that the drug's demonstrated benefit is genuine, not the result of chance. But chance turns out to be a hard thing to rule out. When the measured effects are small — as they are in the vast majority of clinical trials — mere chance is often the difference between whether a drug is deemed to work or not, says John P. A. Ioannidis, a professor of medicine at Stanford.

In a famous 2005 paper published in The Journal of the American Medical Association, Dr. Ioannidis, an authority on statistical analysis, examined nearly four dozen high-profile trials that found a specific medical intervention to be effective. Of the 26 randomized, controlled studies that were followed up by larger trials (examining the same therapy in a bigger pool of patients), the initial finding was wholly contradicted in three cases (12 percent). And in another 6 cases (23 percent), the later trials found the benefit to be less than half of what was first reported.

It wasn't the therapy that changed in each case, but rather the sample size. And Dr. Ioannidis believes that if more rigorous, follow-up studies were actually done, the refutation rate would be far higher.

Donald A. Berry, a professor of biostatistics at M. D. Anderson, agrees. He, too, can rattle off dozens of examples of this evaporation effect and has made a sport, he says, of predicting it. The failures of the last 20 or so Phase 3 trials testing drugs for Alzheimer's disease, he says, could have been predicted based on the lackluster results from Phase 2. Still, the payoff for a successful Phase 3 trial can be so enormous that drug makers will often roll the dice — not on the prospect that the therapy will suddenly work, but on the chance that a trial will suggest that it does.

At a round-table discussion a few years ago, focused on the high failure rate for Alzheimer's drugs, Dr. Berry was amazed to hear one drug company researcher admit to such thinking

out loud. The researcher said that when he and his team designed the Phase 3 trial, he thought the drug would probably fail. But if they could get an approval for a drug for Alzheimer's disease, it would be "a huge success."

"What he was saying," marvels Dr. Berry, "was, 'We're playing the lottery.' "

The fact that the pharmaceutical companies sponsor and run the bulk of investigative drug trials brings what Dr. Ioannidis calls a "constellation of biases" to the process. Too often, he says, trials are against "a straw-man comparator" like a placebo rather than a competing drug. So the studies don't really help us understand which treatments for a disease work best.

But a more fundamental challenge has to do with the nature of clinical trials themselves. "When you do any kind of trial, you're really trying to answer a question about truth in the universe," says Hal Barron, the chief medical officer and head of global development at Roche and Genentech. "And, of course, we can't know that. So we try to design an experiment on a subpopulation of the world that we think is generalizable to the overall universe" — that is, to the patients who will use the drug.

That's a very hard thing to pull off. The rules that govern study enrollment end up creating trial populations that invariably are much younger, have fewer health complications and have been exposed to far less medical treatment than those who are likely to use the drug.

Roughly 53 percent of new cancer diagnoses, for example, are in people 65 or older, but this age group accounts for just 33 percent of participants in cancer drug trials.

Even if clinical researchers could match the demographics of study populations to those of the likely users of these medicines, no group of trial volunteers could ever match the extraordinary biological diversity of the drugs' eventual consumers.

Drug makers are well aware of the challenge. "Listen, it's not lost on anybody that about 95 percent of drugs that enter clinical testing fail to ever get approved," says Dr. Barron. "It's not hard to imagine that at least some of those might have failed because they work very, very well in a small group. We can't continue to have failures due to a lack of appreciation of this heterogeneity in diseases."

So what's the solution? For subtypes of disease that are already known, it may be feasible to design small clinical trials and enroll only those who have the appropriate genetic or molecular signature. That's what Genentech did in developing the breast cancer drug

Herceptin, which homes in on tumor cells that have an abundance of a protein called HER2.

And that's the strategy the company says it's pursuing now. Sixty percent of the new drugs in the works at Genentech/Roche are being developed with a companion diagnostic test to identify the patients who are most likely to benefit.

But given the dismal success rate for drug development, this piecemeal approach is bound to be slow and arduous. Rather than try to fit patients, a handful at a time, into the decades-old clinical-trials framework, we'd be far better off changing the trials themselves.

In fact, a breast cancer trial called I-SPY 2, already under way, may be a good model to follow. The aim of the trial, sponsored by the Biomarkers Consortium, a partnership that includes the Foundation for the National Institutes of Health, the F.D.A., and others, is to figure out whether neoadjuvant therapy for breast cancer — administering drugs before a tumor is surgically removed — reduces recurrence of the disease, and if so, which drugs work best.

As with the Herceptin model, patients are being matched with experimental medicines that are designed to target a particular molecular subtype of breast cancer. But unlike in other trials, I-SPY 2 investigators, including Dr. Berry, are testing up to a dozen drugs from multiple companies, phasing out those that don't appear to be working and subbing in others, without stopping the study.

Part of the novelty lies in a statistical technique called Bayesian analysis that lets doctors quickly glean information about which therapies are working best. There's no certainty in the assessment, but doctors get to learn during the process and then incorporate that knowledge into the ongoing trial.

Mark Gilbert, for his part, would even settle for something simpler in his next glioblastoma study. His definition of a successful clinical trial? "At the end of the day," he says, "regardless of the result, you've learned something."

Clifton Leaf is the author of "The Truth in Small Doses: Why We're Losing the War on Cancer — and How to Win It."