

When Evidence Says No, But Doctors Say Yes

Years after research contradicts common practices, patients continue to demand them and doctors continue to deliver. The result is an epidemic of unnecessary and unhelpful treatment.

by David Epstein, ProPublica

February 22, 2017

This story was co-published with The Atlantic.

FIRST, LISTEN TO THE STORY with the happy ending: At 61, the executive was in excellent health. His blood pressure was a bit high, but everything else looked good, and he exercised regularly. Then he had a scare. He went for a brisk post-lunch walk on a cool winter day, and his chest began to hurt. Back inside his office, he sat down, and the pain disappeared as quickly as it had come.

That night, he thought more about it: middle-aged man, high blood pressure, stressful job, chest discomfort. The next day, he went to a local emergency department. Doctors determined that the man had not suffered a heart attack and that the electrical activity of his heart was completely normal. All signs suggested that the executive had stable angina — chest pain that occurs when the heart muscle is getting less blood-borne oxygen than it needs, often because an artery is partially blocked.

A cardiologist recommended that the man immediately have a coronary angiogram, in which a catheter is threaded into an artery to the heart and injects a dye that then shows up on special x-rays that look for blockages. If the test found a blockage, the cardiologist advised, the executive should get a stent, a metal tube that slips into the artery and forces it open.

While he was waiting in the emergency department, the executive took out his phone and searched “treatment of coronary artery disease.” He immediately found information from medical journals that said medications, like aspirin and blood-pressure-lowering drugs, should be the first line of treatment. The man was an unusually self-possessed patient, so he asked the cardiologist about what he had found. The cardiologist was dismissive and told the man to “do more research.” Unsatisfied, the man declined to have the angiogram and consulted his primary-care doctor.

The primary-care physician suggested a different kind of angiogram, one that did not require a catheter but instead used multiple x-rays to image arteries. That test revealed an artery that was partially blocked by plaque, and though the man’s heart was pumping blood normally, the test was incapable of determining whether the blockage was dangerous. Still, his primary-care doctor, like the cardiologist at the emergency room, suggested that the executive have an angiogram with a catheter, likely followed by a procedure to implant a stent. The man set up an appointment with the cardiologist he was referred to for the catheterization, but when he tried to contact that

doctor directly ahead of time, he was told the doctor wouldn't be available prior to the procedure. And so the executive sought yet another opinion. That's when he found Dr. David L. Brown, a professor in the cardiovascular division of the Washington University School of Medicine in St. Louis. The executive told Brown that he'd felt pressured by the previous doctors and wanted more information. He was willing to try all manner of noninvasive treatments — from a strict diet to retiring from his stressful job — before having a stent implanted.

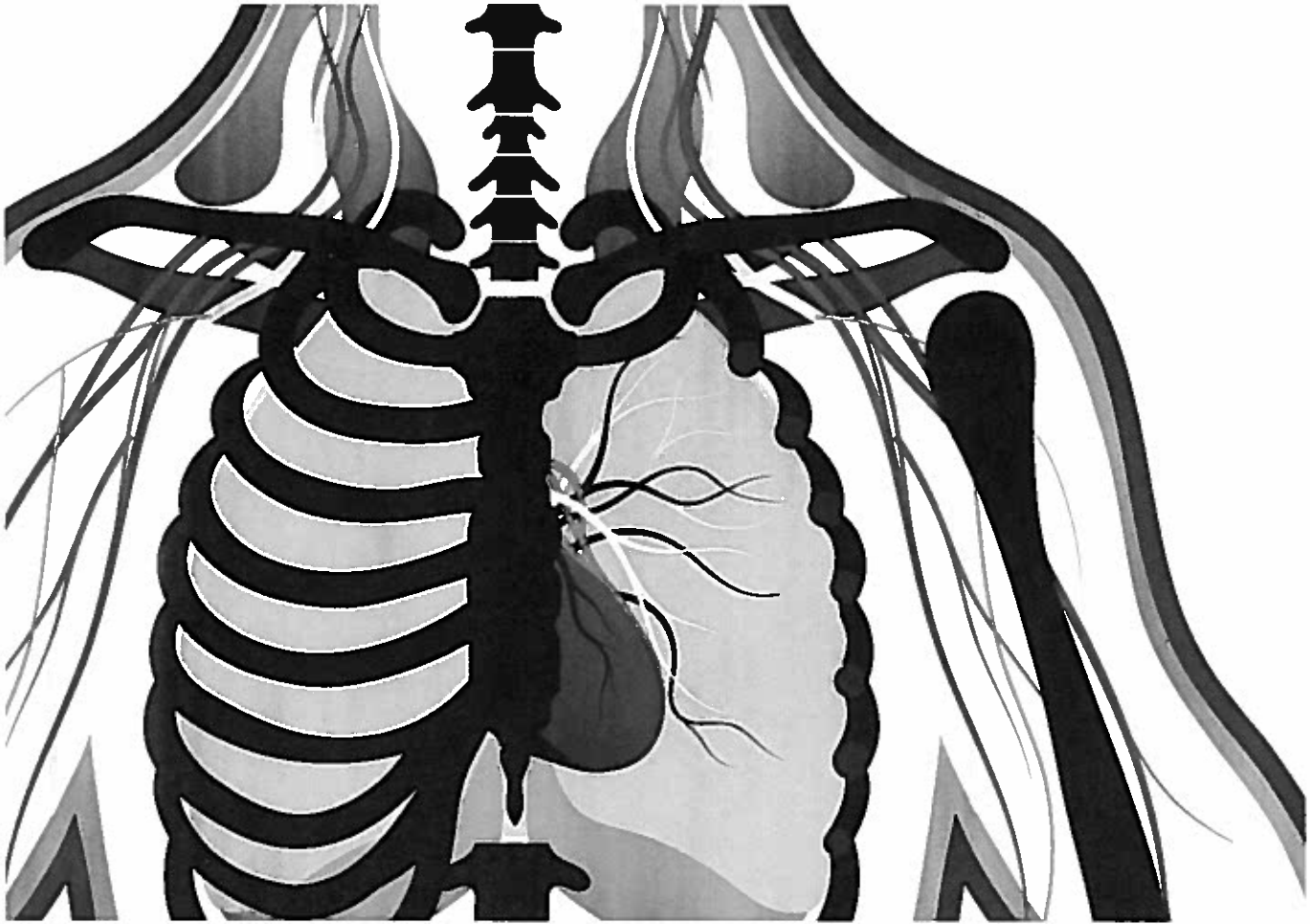
The executive had been very smart to seek more information, and now, by coming to Brown, he was very lucky, too. Brown is part of the RightCare Alliance, a collaboration between health-care professionals and community groups that seeks to counter a trend: increasing medical costs without increasing patient benefits. As Brown put it, RightCare is "bringing medicine back into balance, where everybody gets the treatment they need, and nobody gets the treatment they don't need." And the stent procedure was a classic example of the latter. In 2012, Brown had coauthored a paper that examined every randomized clinical trial that compared stent implantation with more conservative forms of treatment, and he found that stents for stable patients prevent zero heart attacks and extend the lives of patients a grand total of not at all. In general, Brown says, "nobody that's not having a heart attack needs a stent." (Brown added that stents may improve chest pain in some patients, albeit fleetingly.) Nonetheless, hundreds of thousands of stable patients receive stents annually, and one in 50 will suffer a serious complication or die as a result of the implantation procedure.

Brown explained to the executive that his blockage was one part of a broader, more diffuse condition that would be unaffected by opening a single pipe. The cardiovascular system, it turns out, is more complicated than a kitchen sink. The executive started medication and improved his diet. Three months later, his cholesterol had improved markedly, he had lost 15 pounds, and the chest pain never returned.

Now, listen to the story with the sad ending: Not long after helping the executive, Brown and his colleagues were asked to consult on the case of a 51-year-old man from a tiny Missouri town. This man had successfully recovered from Hodgkin's lymphoma, but radiation and six cycles of chemotherapy had left him with progressive scarring creeping over his lungs. He was suffocating inside his own body. The man was transferred to Barnes Jewish Hospital, where Brown works, for a life-saving lung transplant. But when the man arrived in St. Louis, the lung-transplant team could not operate on him.

Four months earlier, the man had been admitted to another hospital because he was having trouble breathing. There, despite the man's history of lymphoma treatment, which can cause scarring, a cardiologist wondered whether the shortness of breath might be due to a blocked artery. As with the executive, the cardiologist recommended a catheter. Unlike the executive, however, this man, like most patients, agreed to the procedure. It revealed a partial blockage of one coronary artery. So, doctors implanted a stent, even though there was no clear evidence that the blockage was responsible for the man's shortness of breath — which was, in fact, caused by the lung scarring. Finally, the man was put on standard post-implantation medications to make sure he would not develop a blood clot at the site of the stent. But those medications made surgery potentially lethal, putting the man at an extremely high risk of bleeding to death during the transplant. The operation had to be delayed.

Meanwhile, the man's lung tissue continued to harden and scar, like molten lava that cools and hardens into gray stone. Until one day, he couldn't suck in another breath. The man had survived advanced-stage lymphoma only to die in the hospital, waiting until he could go off needed medication for an unneeded stent.



WHAT THE PATIENTS IN BOTH STORIES had in common was that neither needed a stent. By dint of an inquiring mind and a smartphone, one escaped with his life intact. The greater concern is: How can a procedure so contraindicated by research be so common?

When you visit a doctor, you probably assume the treatment you receive is backed by evidence from medical research. Surely, the drug you're prescribed or the surgery you'll undergo wouldn't be so common if it didn't work, right?

For all the truly wondrous developments of modern medicine — imaging technologies that enable precision surgery, routine organ transplants, care that transforms premature infants into perfectly healthy kids, and remarkable chemotherapy treatments, to name a few — it is distressingly ordinary for patients to get treatments that research has shown are ineffective or even dangerous. Sometimes doctors simply haven't kept up with the science. Other times doctors know the state of play perfectly well but continue to deliver these treatments because it's profitable — or even because they're popular and patients demand them. Some procedures are implemented based on studies that did not prove whether they really worked in the first place. Others were initially supported by evidence but then were contradicted by better evidence, and yet these procedures have remained the standards of care for years, or decades.

Even if a drug you take was studied in thousands of people and shown truly to save lives, chances are it won't do that for you. The good news is, it probably won't harm you, either. Some of the most widely prescribed medications do little of anything meaningful, good or bad, for most people who take them.

In a 2013 study, a dozen doctors from around the country examined all 363 articles published in *The New England Journal of Medicine* over a decade — 2001 through 2010 — that tested a current clinical practice, from the use of antibiotics to treat people with persistent Lyme disease symptoms (didn't help) to the use of specialized sponges for

preventing infections in patients having colorectal surgery (caused more infections). Their results, published in the Mayo Clinic Proceedings, found 146 studies that proved or strongly suggested that a current standard practice either had no benefit at all or was inferior to the practice it replaced; 138 articles supported the efficacy of an existing practice, and the remaining 79 were deemed inconclusive. (There was, naturally, plenty of disagreement with the authors' conclusions.) Some of the contradicted practices possibly affect millions of people daily: Intensive medication to keep blood pressure very low in diabetic patients caused more side effects and was no better at preventing heart attacks or death than more mild treatments that allowed for a somewhat higher blood pressure. Other practices challenged by the study are less common — like the use of a genetic test to determine if a popular blood thinner is right for a particular patient — but gaining in popularity despite mounting contrary evidence. Some examples defy intuition: CPR is no more effective with rescue breathing than if chest compressions are used alone; and breast-cancer survivors who are told not to lift weights with swollen limbs actually should lift weights, because it improves their symptoms.

A separate but similarly themed study in 2012 funded by the Australian Department of Health and Ageing, which sought to reduce spending on needless procedures, looked across the same decade and identified 156 active medical practices that are probably unsafe or ineffective. The list goes on: A brand new review of 48 separate studies — comprising more than 13,000 clinicians — looked at how doctors perceive disease-screening tests and found that they tend to underestimate the potential harms of screening and overestimate the potential benefits; an editorial in *American Family Physician*, co-written by one of the journal's editors, noted that a "striking feature" of recent research is how much of it contradicts traditional medical opinion.

That isn't likely to change any time soon. The 21st Century Cures Act — a rare bipartisan bill, pushed by more than 1,400 lobbyists and signed into law in December — lowers evidentiary standards for new uses of drugs and for marketing and approval of some medical devices. Furthermore, last month President Donald Trump scolded the FDA for what he characterized as withholding drugs from dying patients. He promised to slash regulations "big league. ... It could even be up to 80 percent" of current FDA regulations, he said. To that end, one of the president's top candidates to head the FDA, tech investor Jim O'Neill, has openly advocated for drugs to be approved before they're shown to work. "Let people start using them at their own risk," O'Neill has argued.

Stents for stable patients prevent zero heart attacks and extend the lives of patients a grand total of not at all.

So, while Americans can expect to see more drugs and devices sped to those who need them, they should also expect the problem of therapies based on flimsy evidence to accelerate. In a recent *Stat* op-ed, two Johns Hopkins University physician-researchers wrote that the new 21st Century Cures Act will turn the label "FDA approved" into "a shadow of its former self." In 1962, Congress famously raised the evidentiary bar for drug approvals after thousands of babies were born with malformed limbs to mothers who had taken the sleep aid thalidomide. Steven Galson, a retired rear admiral and former acting surgeon general under both President George W. Bush and President Barack Obama, has called the strengthened approval process created in 1962 the FDA's "biggest contribution to health." Before that, he said, "many marketed drugs were ineffective for their labeled uses."

Striking the right balance between innovation and regulation is incredibly difficult, but once remedies are in use — even in the face of contrary evidence — they tend to persist. A 2007 *Journal of the American Medical Association* paper coauthored by John Ioannidis — a Stanford University medical researcher and statistician who rose to

prominence exposing poor-quality medical science — found that it took 10 years for large swaths of the medical community to stop referencing popular practices after their efficacy was unequivocally vanquished by science.

According to Vinay Prasad, an oncologist and one of the authors of the Mayo Clinic Proceedings paper, medicine is quick to adopt practices based on shaky evidence but slow to drop them once they've been blown up by solid proof. As a young doctor, Prasad had an experience that left him determined to banish ineffective procedures. He was the medical resident on a team caring for a middle-aged woman with stable chest pain. She underwent a stent procedure and suffered a stroke, resulting in brain damage. Prasad, now at the Oregon Health and Sciences University, still winces slightly when he talks about it. University of Chicago professor and physician Adam Cifu had a similar experience. Cifu had spent several years convincing newly postmenopausal patients to go on hormone therapy for heart health — a treatment that at the millennium accounted for 90 million annual prescriptions — only to then see a well-designed trial show no heart benefit and perhaps even a risk of harm. "I had to basically run back all those decisions with women," he says. "And, boy, that really sticks with you, when you have patients saying, 'But I thought you said this was the right thing.'" So he and Prasad coauthored a 2015 book, "Ending Medical Reversal," a call to raise the evidence bar for adopting new medical standards. "We have a culture where we reward discovery; we don't reward replication," Prasad says, referring to the process of retesting initial scientific findings to make sure they're valid.

Steven Nissen, chairman of cardiovascular medicine at the Cleveland Clinic, says the situation with stents, at least, is improving. As a previous president of the American College of Cardiology, he helped create guidelines for determining when a stable patient might be a reasonable candidate for a stent. (Both Nissen and David Holmes, a Mayo Clinic cardiologist and also a former ACC president, said that in cases in which patients have had bad responses to medication and persistent, life-altering chest pain, even a short-term reduction of symptoms may justify a stent.) Thanks to such guidelines, the frequency of clearly inappropriate stent placement declined significantly between 2010 and 2014. Still, the latest assessment in more than 1,600 hospitals across the country concluded that about half of all stent placements in stable patients were either definitely or possibly inappropriate. "Things have gotten better," Nissen says, "but they're not where they need to be." Nissen thinks removing financial incentives can also help change behavior. "I have a dozen or so cardiologists, and they get the exact same salary whether they put in a stent or don't," Nissen says, "and I think that's made a difference and kept our rates of unnecessary procedures low."

Two years ago, a trio of Bloomberg journalists reported that Mount Sinai Hospital in New York City was scheduling "emergencies-by-appointment" for patients to get stents, because, the report said, insurance is more likely to cover the procedure in an emergency situation. (For a patient who is having a heart attack, a stent can be life-saving.) Mount Sinai's catheter lab features annual reports that boast of how many stents are implanted, alongside patient testimonials, like one from 77-year-old Nelly Rodriguez, who notes that her doctor "reassures me that as long as I follow his instructions, eat healthy, and remain smoke-free, the stents he has put into my arteries over the years should last and I will feel well." In most cases, every word of that sentence between "smoke-free" and "I will feel well" could be deleted and it would be just as true.

It is, of course, hard to get people in any profession to do the right thing when they're paid to do the wrong thing. But there's more to this than market perversion. On a recent snowy St. Louis morning, Brown gave a grand-rounds lecture to about 80 doctors at Barnes Jewish Hospital. Early in the talk, he showed results from medical tests on the executive he treated, the one who avoided a stent. He then presented data from thousands of patients in randomized controlled trials of stents versus noninvasive treatments, and it showed that stents yielded no benefit for stable patients. He asked the doctors in the room to raise their hands if they would still send a patient with the same diagnostic findings as the executive for a catheterization, which would almost surely lead to a stent. At least half of the hands in the room went up, some of them sheepishly. Brown expressed surprise at the honesty in the room. "Well," one of the attendees told him, "we know what we do." But why?

In 2007, after a seminal study, the COURAGE trial, showed that stents did not prevent heart attacks or death in stable patients, a trio of doctors at the University of California, San Francisco, conducted 90-minute focus groups with cardiologists to answer that question. They presented the cardiologists with fictional scenarios of patients who had at least one narrowed artery but no symptoms and asked them if they would recommend a stent. Almost to a person, the cardiologists, including those whose incomes were not tied to tests and procedures, gave the same answers: They said that they were aware of the data but would still send the patient for a stent. The rationalizations in each focus group followed four themes: (1) Cardiologists recalled stories of people dying suddenly — including the highly publicized case of jogging guru Jim Fixx — and feared they would regret it if a patient did not get a stent and then dropped dead. The study authors concluded that cardiologists were being influenced by the “availability heuristic,” a term coined by Nobel laureate psychologists Amos Tversky and Daniel Kahneman for the human instinct to base an important decision on an easily recalled, dramatic example, even if that example is irrelevant or incredibly rare. (2) Cardiologists believed that a stent would relieve patient anxiety. (3) Cardiologists felt they could better defend themselves in a lawsuit if a patient did get a stent and then died, rather than if they didn’t get a stent and died. “In California,” one said, “if this person had an event within two years, the doctor who didn’t intervene would be successfully sued.” And there was one more powerful and ubiquitous reason: (4) Despite the data, cardiologists couldn’t believe that stents did not help: Stenting just made so much sense. A patient has chest pain, a doctor sees a blockage, how can opening the blockage not make a difference?

In the late 1980s, with evidence already mounting that forcing open blood vessels was less effective and more dangerous than noninvasive treatments, cardiologist Eric Topol coined the term, “oculostenotic reflex.” Oculo, from the Latin for “eye,” and stenotic, from the Greek for “narrow,” as in a narrowed artery. The meaning: If you see a blockage, you’ll reflexively fix a blockage. Topol described “what appears to be an irresistible temptation among some invasive cardiologists” to place a stent any time they see a narrowed artery, evidence from thousands of patients in randomized trials be damned. Stenting is what scientists call “bio-plausible” — intuition suggests it should work. It’s just that the human body is a little more Book of Job and a little less household plumbing: Humans didn’t invent it, it’s really complicated, and people often have remarkably little insight into cause and effect.

CHANCES ARE, you or someone in your family has taken medication or undergone a procedure that is bio-plausible but does not work.

According to the Centers for Disease Control and Prevention, about one in three American adults have high blood pressure. Blood pressure is a measure of how hard your blood is pushing on the sides of vessels as it moves through your body; the harder the pushing, the more strain on your heart. People with high blood pressure are at enormously increased risk for heart disease (the nation’s No. 1 killer) and stroke (No. 3).

So it’s not hard to understand why Sir James Black won a Nobel Prize largely for his 1960s discovery of beta-blockers, which slow the heart rate and reduce blood pressure. The Nobel committee lauded the discovery as the “greatest breakthrough when it comes to pharmaceuticals against heart illness since the discovery of digitalis 200 years ago.” In 1981, the FDA approved one of the first beta-blockers, atenolol, after it was shown to dramatically lower blood pressure. Atenolol became such a standard treatment that it was used as a reference drug for comparison with other blood-pressure drugs.

In 1997, a Swedish hospital began a trial of more than 9,000 patients with high blood pressure who were randomly assigned to take either atenolol or a competitor drug that was designed to lower blood pressure for at least four years. The competitor-drug group had fewer deaths (204) than the atenolol group (234) and fewer strokes (232 compared with 309). But the study also found that both drugs lowered blood pressure by

the exact same amount, so why wasn't the vaunted atenolol saving more people? That odd result prompted a subsequent study, which compared atenolol with sugar pills. It found that atenolol didn't prevent heart attacks or extend life at all; it just lowered blood pressure. A 2004 analysis of clinical trials — including eight randomized controlled trials comprising more than 24,000 patients — concluded that atenolol did not reduce heart attacks or deaths compared with using no treatment whatsoever; patients on atenolol just had better blood-pressure numbers when they died.

“Yes, we can move a number, but that doesn't necessarily translate to better outcomes,” says John Mandrola, a cardiac electrophysiologist in Louisville who advocates for healthy lifestyle changes. It's tough, he says, “when patients take a pill, see their numbers improve, and think their health is improved.”

The overall picture of beta-blockers is complex. For example, some beta-blockers have been shown clearly to reduce the chance of a stroke or heart attack in patients with heart failure. But the latest review of beta-blockers from the Cochrane Collaboration — an independent, international group of researchers that attempts to synthesize the best available research — reported that they “are not recommended as first line treatment for hypertension as compared to placebo due to their modest effect on stroke and no significant reduction in mortality or coronary heart disease.”

Researchers writing in *Lancet* questioned the use of atenolol as a comparison standard for other drugs and added that “stroke was also more frequent with atenolol treatment” compared with other therapies. Still, according to a 2012 study in the *Journal of the American Medical Association*, more than 33.8 million prescriptions of atenolol were written at a retail cost of more than \$260 million. There is some evidence that atenolol might reduce the risk of stroke in young patients, but there is also evidence that it increases the risk of stroke in older patients — and it is older patients who are getting it en masse. According to ProPublica's Medicare prescription database, in 2014, atenolol was prescribed to more than 2.6 million Medicare beneficiaries, ranking it the 31st most prescribed drug out of 3,362 drugs. One doctor, Chinh Huynh, a family practitioner in Westminster, California, wrote more than 1,100 atenolol prescriptions in 2014 for patients over 65, making him one of the most prolific prescribers in the country. Reached at his office, Huynh said atenolol is “very common for hypertension; it's not just me.” When asked why he continues to prescribe atenolol so frequently in light of the randomized, controlled trials that showed its ineffectiveness, Huynh said, “I read a lot of medical magazines, but I didn't see that.” Huynh added that his “patients are doing fine with it” and asked that any relevant journal articles be faxed to him.

Brown, the Washington University cardiologist, says that once doctors get out of training, “it's a job, and they're trying to earn money, and they don't necessarily keep up. So really major changes have to be generational.”

Data compiled by QuintilesIMS, which provides information and technology services to the health-care industry, show that atenolol prescriptions consistently fell by 3 million per year over a recent five-year period. If that rate holds, atenolol will stop being prescribed in just under two decades since high-quality trials showed that it simply does not work.

Atenolol did not reduce any more heart attacks or deaths than using no treatment whatsoever; patients on atenolol just had better blood-pressure numbers when they died.

JUST AS THE CARDIOVASCULAR SYSTEM is not a kitchen sink, the musculoskeletal system is not an erector set. Cause and effect is frequently elusive.

Consider the knee, that most bedeviling of joints. A procedure known as arthroscopic partial meniscectomy, or APM, accounts for roughly a half-million procedures per year at a cost of around \$4 billion. A meniscus is a crescent-shaped piece of fibrous cartilage that helps stabilize and provide cushioning for the knee joint. As people age, they often suffer tears in the meniscus that are not from any acute injury. APM is meant to relieve knee pain by cleaning out damaged pieces of a meniscus and shaving the cartilage back to crescent form. This is not a fringe surgery; in recent years, it has been one of the most popular surgical procedures in the hemisphere. And a burgeoning body of evidence says that it does not work for the most common varieties of knee pain.

Something like the knee version of the oculostenotic reflex takes hold: A patient comes in with knee pain, and an MRI shows a torn meniscus; naturally, the patient wants it fixed, and the surgeon wants to fix it and send the patient for physical therapy. And patients do get better, just not necessarily from the surgery.

A 2013 study of patients over 45 conducted in seven hospitals in the United States found that APM followed by physical therapy produced the same results as physical therapy alone for most patients. Another study at two public hospitals and two physical-therapy clinics found the same result two years after treatment.

A unique study at five orthopedic clinics in Finland compared APM with “sham surgery.” That is, surgeons took patients with knee pain to operating rooms, made incisions, faked surgeries, and then sewed them back up. Neither the patients nor the doctors evaluating them knew who had received real surgeries and who was sporting a souvenir scar. A year later, there was nothing to tell them apart. The sham surgery performed just as well as real surgery. Except that, in the long run, the real surgery may increase the risk of knee osteoarthritis. Also, it’s expensive, and, while APM is exceedingly safe, surgery plus physical therapy has a greater risk of side effects than just physical therapy.

At least one-third of adults over 50 will show meniscal tears if they get an MRI. But two-thirds of those will have no symptoms whatsoever. (For those who do have pain, it may be from osteoarthritis, not the meniscus tear.) They would never know they had a tear if not for medical imaging, but once they have the imaging, they may well end up having surgery that doesn’t work for a problem they don’t have.

For obvious reasons, placebo-controlled trials of surgeries are difficult to execute. The most important question then is: Why, when the highest level of evidence available contradicts a common practice, does little change?

For one, the results of these studies do not prove that the surgery is useless, but rather that it is performed on a huge number of people who are unlikely to get any benefit. Meniscal tears are as diverse as the human beings they belong to, and even large studies will never capture all the variation that surgeons see; there are compelling real-world results that show the surgery helps certain patients. “I think it’s an extremely helpful intervention in cases where a patient does not suffer from the constant ache of arthritis, but has sharp, intermittent pain and a blockage of motion,” says John Christoforetti, a prominent orthopedic surgeon in Pittsburgh. “But when you’re talking about the average inactive American, who suffers gradual onset knee pain and has full motion, many of them have a meniscal tear on MRI and they should not have surgery as initial treatment.”

Still, the surgery — like some others meant for narrower uses — is common even for patients who don't need it. And patients themselves are part of the problem. According to interviews with surgeons, many patients they see want, or even demand, to be operated upon and will simply shop around until they find a willing doctor. Christoforetti recalls one patient who traveled a long way to see him but was “absolutely not a candidate for an operation.” Despite the financial incentive to operate, he explained to the patient and her husband that the surgery would not help. “She left with a smile on her face,” Christoforetti says, “but literally as they're checking out, we got a ding that someone had rated us [on a website], and it's her husband. He's been typing on his phone during the visit, and it's a one-star rating that I'm this insensitive guy he wouldn't let operate on his dog. They'd been online, and they firmly believed she needed this one operation and I was the guy to do it.”

So, what do surgeons do? “Most of my colleagues,” Christoforetti says, “will say: ‘Look, save yourself the headache, just do the surgery. None of us are going to be upset with you for doing the surgery. Your bank account's not going to be upset with you for doing the surgery. Just do the surgery.’”

RANDOMIZED, PLACEBO-CONTROLLED TRIALS are the gold standard of medical evidence. But not all RCTs, as they are known, are created equal. Even within the gold standard, well-intentioned practices can muddle a study. That is particularly true with “crossover” trials, which have become popular for cancer-drug investigations.

In cancer research, a crossover trial often means that patients in the control group, who start on a placebo, are actually given the experimental drug during the study if their disease progresses. Thus, they are no longer a true control group. The benefit of a crossover trial is that it allows more people with severe disease to try an experimental drug; the disadvantage is the possibility that the study is altered in a manner that obscures the efficacy of the drug being tested.

In 2010, on the strength of a crossover trial, Provenge became the first cancer vaccine approved by the FDA. A cancer vaccine is a form of immunotherapy, in which a patient's own immune system is spurred by a drug to attack cancer cells. Given the extraordinary difficulty of treating metastatic cancer, and high expectations following the abject failure of other cancer vaccines, the approval of Provenge was greeted with ecstatic enthusiasm. One scientific paper heralded it as “the gateway to an exciting new paradigm.” Except, Provenge did not hinder tumor growth at all, and it's hard to know if it really works.

Provenge was approved based on the “IMPACT study,” a randomized, placebo-controlled trial initially meant to see whether Provenge could stop prostate cancer from progressing. It didn't. Three-and-a-half months into the study, the cancers of patients who had received Provenge and those who had received a placebo had advanced similarly. Nonetheless, patients who received Provenge ultimately had a median survival time of about four months longer than those who received the placebo. Due to the way in which the IMPACT trial unfolded, however, it's hard to tell if Provenge was truly responsible for the life extension.

Because Provenge did not halt tumor growth, many of the patients who began the study on it also started to receive docetaxel, a chemotherapy drug that is well established to treat advanced prostate cancer. The cancers of the patients on a placebo were also progressing, so they were “crossed over” and given Provenge after a delay. Their cancer continued progressing, and after another delay, many of them also got docetaxel. In the end, fewer patients in the group that started on a placebo received docetaxel, and, when they did, they got it later in the study. So Provenge may have worked, but it's impossible to tell for sure: Was the slightly longer survival of one group because they got Provenge earlier or because the other group got docetaxel later?

The year after Provenge was approved, the federal government’s Agency for Healthcare Research and Quality issued a “technology assessment” report examining all of the evidence regarding Provenge efficacy. The report says there is “moderate” evidence that Provenge effectively treats cancer, but it also highlighted the fact that more patients who got Provenge at the beginning of the seminal trial also received more and earlier chemotherapy. The report concludes that the effect of Provenge is apparent “only in the context of a substantial amount of eventual chemotherapeutic treatment.” In other words, it is unclear which effects in the trial were due to Provenge and which were due to chemotherapy.

“The people who went on docetaxel went on it because their disease was progressing, so you’ve already broken the randomization,” says Elise Berliner, director of AHRQ’s Technology Assessment Program. Prasad, the oncologist who advocates for higher standards of preapproval evidence, is less diplomatic: “If the treatment were Pixy Stix, you’d have a similar effect. One group gets Pixy Stix, and when their cancer progresses, they get a real treatment.”

The larger issue has nothing to do with Provenge specifically but about the way it gained FDA approval. Therapies are frequently approved for use based on clinical trials that can’t actually prove whether they work. “Clinical trials almost all have issues like this one,” Berliner says, “and it’s very hard to do randomized controlled trials after drugs are approved.” According to a new paper in the *Journal of the American Medical Association Oncology*, even when cancer drugs clearly do work in trials, they often don’t work or work substantially less well in the real world, perhaps because subjects in trials are not representative of typical patients. Berliner is hoping to expand and improve registries that track large numbers of real-world patients as an additional source of information. “I’ve been here for 15 years producing these reports,” she says, “and I’m getting frustrated.”

“Just do the surgery. None of us are going to be upset with you for doing the surgery. Your bank account’s not going to be upset with you for doing the surgery. Just do the surgery.”

IDEALLY, FINDINGS THAT SUGGEST a therapy works and those that suggest it does not would receive attention commensurate with their scientific rigor, even in the earliest stages of exploration. But academic journals, scientists, and the media all tend to prefer research that concludes that some exciting new treatment does indeed work.

In 2012, a team of scientists from UCLA published an article in the prominent *New England Journal of Medicine*, the most cited medical journal in the world, showing that deep brain stimulation — delivered via electrodes implanted in the brains of Parkinson’s patients — improved spatial memory, a lot. The study was understandably small — just seven subjects — as there are only so many people with electrodes already implanted in their brains. It was covered in outlets like *The New York Times* (“Study Explores Electrical Stimulation as An Aid to Memory”), *The Wall Street Journal* (“Memory Gets Jolt in Brain Research”), and *LiveScience* (“Where Did I Park? Brain Treatment May Enhance Spatial Memory”). The *NEJM* itself published an editorial in the same issue noting that the study was “preliminary, is based on small samples, and requires replication” but was worth following up with “well-designed studies.”

Given the potential impact, an international team led by Joshua Jacobs, a biomedical-engineering professor at Columbia University, set out to replicate the initial finding with a larger sample. “If it did indeed work, it would be a very important approach that could help people,” Jacobs says. The team took several years and tested 49 subjects, so that their study would give more statistically reliable results. The scientists were rather stunned to find that deep brain stimulation actually impaired spatial memory in their study. It was a disappointing result, but they were encouraged to show that brain stimulation could affect memory at all — a step toward figuring out how to wield such technology — and they felt an obligation to submit it to the NEJM. That is how science is supposed to work, after all, because failing to publish negative results is recognized to be a massive source of scientific misinformation.



REPLICATION OF RESULTS in science was a cause-célèbre last year, due to the growing realization that researchers have been unable to duplicate a lot of high-profile results. A decade ago, Stanford’s Ioannidis published a paper warning the scientific community that “Most Published Research Findings Are False.” (In 2012, he coauthored a paper showing that pretty much everything in your fridge has been found to both cause and prevent cancer — except bacon, which apparently only causes cancer.) Ioannidis’s prescience led his paper to be cited in other scientific articles more than 800 times in 2016 alone. Point being, sensitivity in the scientific community to replication problems is at an all-time high. So Jacobs and his coauthors were bemused when the NEJM rejected their paper.

One of the reviewers (peer reviewers are anonymous) who rejected the paper gave this feedback: “Much more interesting would have been to find a set of stimulation parameters that would enhance memory.” In other words: The paper would be better if, like the original study, it had found a positive rather than a negative result. (Last spring, ProPublica wrote about heavy criticism of the NEJM’s reluctance to publish research that questioned earlier findings.)

Another reviewer noted that electrodes were placed on most of the subjects differently in the replication study compared with those in the original study. So Jacobs and his coauthors analyzed results only from patients with the exact same electrode placement as the original study, and the findings were the same. Three of the authors wrote back to the NEJM, pointing out errors in the reviewer comments; they received a short note back saying that the paper rejection “was not based on the specific comments of the reviewers you discuss in your response letter” and that the journal gets many more papers than it can print. That is, of course, very true, particularly for important journals. *Neuron*, one of the most prominent neuroscience-specific journals, quickly accepted the paper and published it last month. (It did not receive the media fanfare of the original paper — or almost any at all — although *The Wall Street Journal* did cover it.)

The same week the paper appeared in *Neuron*, Columbia University held a daylong symposium to discuss the replication problem in science. The president of the National Academy of Sciences and the director of the U.S. Office of Research Integrity spoke — so too did Jeffrey Drazen, editor-in-chief of the NEJM. Jacobs was in the audience.

In the final Q&A, Jacobs stepped up to one of the audience microphones and asked Drazen if journals had an obligation to publish high-quality replication attempts of prominent studies, and he disclosed that his team’s had been rejected by the NEJM. Drazen declined to discuss Jacobs’s paper, but he said that “as editors, we’re powerless,” and the onus should be on the replication researchers, or “the complainant,” as he put it, “and the original paper author to work together toward the truth. We’re not trying to say who’s right and who’s wrong; we’re trying to find out what we need to know. Veritas, to advance human health, it’s that simple.”

Jacobs did not find the answer that simple. He found it strange. On a panel about transparency and replication, Drazen seemed to be saying that journals, the main method of information dissemination and the primary forum for replication in science, could do little and that “complainants” need to sort it out with *de facto* defendants. Many doctors, scientists, patient advocates, and science writers keep track of new developments through premier publications like the NEJM. The less publicly a shaky scientific finding is challenged, the more likely it becomes entrenched common knowledge.

OF COURSE, MYRIAD MEDICAL INNOVATIONS improve and save lives, but even as scientists push the cutting edge (and expense) of medicine, the National Center for Health Statistics reported last month that American life expectancy dropped, slightly. There is, though, something that does powerfully and assuredly bolster life expectancy: sustained public-health initiatives.

Medicine can be like wine: Expense is sometimes a false signal of quality. On an epochal scale, even the greatest triumphs of modern medicine, like the polio vaccine, had a small impact on human health compared with the impact of better techniques for sanitation and food preservation. Due to smoking and poor lifestyle habits, lung cancer — which killed almost no Americans in the early 20th century — is today by far the biggest killer among cancers. Thankfully, public pressure to curb smoking has put lung-cancer deaths in rapid decline since a peak in the 1990s. Deaths from lung cancer should continue to diminish, as they are tightly correlated to smoking rates — but with a 20-year lag; that is, lung cancer deaths will decline 20 years after smoking rates decline.

The health problems that most commonly afflict the American public are largely driven by lifestyle habits—smoking, poor nutrition, and lack of physical activity, among others. In November, a team led by researchers at Massachusetts General Hospital pooled data from tens of thousands of people in four separate health studies from 1987 to 2008. They found that simple, moderate lifestyle changes dramatically reduced the risk of heart disease, the most prolific killer in the country, responsible for one in every four deaths. People deemed at high familial risk of heart disease cut their risk in half if they satisfied three of the following four criteria: didn’t smoke (even if they smoked in the past); weren’t obese (although they could be overweight); exercised once a week; ate more real food and less processed food. Fitting even two of those categories still substantially decreased risk. In August, a report issued by the

International Agency for Research on Cancer concluded that obesity is now linked to an extraordinary variety of cancers, from thyroids and ovaries to livers and colons.

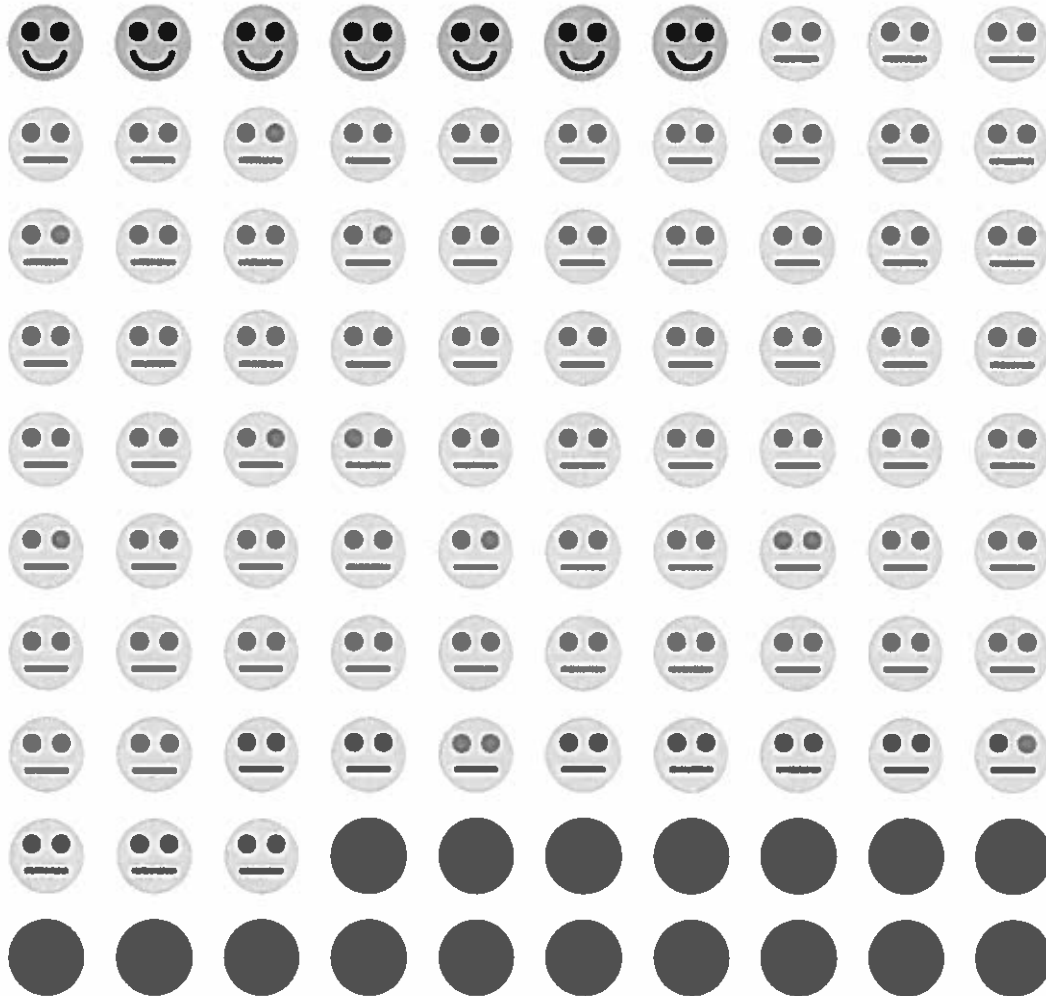
“Relative risk is just another way of lying.”

AT THE SAME TIME, patients and even doctors themselves are sometimes unsure of just how effective common treatments are, or how to appropriately measure and express such things. Graham Walker, an emergency physician in San Francisco, co-runs a website staffed by doctor volunteers called the NNT that helps doctors and patients understand how impactful drugs are — and often are not. “NNT” is an abbreviation for “number needed to treat,” as in: How many patients need to be treated with a drug or procedure for one patient to get the hoped-for benefit? In almost all popular media, the effects of a drug are reported by relative risk reduction. To use a fictional illness, for example, say you hear on the radio that a drug reduces your risk of dying from Hogwart’s disease by 20 percent, which sounds pretty good. Except, that means if 10 in 1,000 people who get Hogwart’s disease normally die from it, and every single patient goes on the drug, eight in 1,000 will die from Hogwart’s disease. So, for every 500 patients who get the drug, one will be spared death by Hogwart’s disease. Hence, the NNT is 500. That might sound fine, but if the drug’s “NNH” — “number needed to harm” — is, say, 20 and the unwanted side effect is severe, then 25 patients suffer serious harm for each one who is saved. Suddenly, the trade-off looks grim.

Now, consider a real and familiar drug: aspirin. For elderly women who take it daily for a year to prevent a first heart attack, aspirin has an estimated NNT of 872 and an NNH of 436. That means if 1,000 elderly women take aspirin daily for a decade, 11 of them will avoid a heart attack; meanwhile, twice that many will suffer a major gastrointestinal bleeding event that would not have occurred if they hadn’t been taking aspirin. As with most drugs, though, aspirin will not cause anything particularly good or bad for the vast majority of people who take it. That is the theme of the medicine in your cabinet: It likely isn’t significantly harming or helping you. “Most people struggle with the idea that medicine is all about probability,” says Aron Sousa, an internist and senior associate dean at Michigan State University’s medical school. As to the more common metric, relative risk, “it’s horrible,” Sousa says. “It’s not just drug companies that use it; physicians use it, too. They want their work to look more useful, and they genuinely think patients need to take this [drug], and relative risk is more compelling than NNT. Relative risk is just another way of lying.”

A Different Way to Think About Medicine

For every 100 older adults who take a sleep aid, 7 will experience improved sleep, while 17 will suffer side effects that range widely in severity, from simple morning “hangover” to memory loss and serious accidents. As with many medications, most who take a sleep aid will experience neither benefit nor harm.



😊 = Effective 😐 = No Benefit or Harm ● = Harmful

Source: Adapted from a Waitemata District Health Board handout, based on data from Glass et. al. BMJ (2005) "Sedative hypnotics in older people with insomnia: meta-analysis of risks and benefits"

Even remedies that work extraordinarily well can be less impressive when viewed via NNT. Antibiotics for a sinus infection will resolve symptoms faster in one of 15 people who get them, while one in eight will experience side effects. A meta-analysis of sleep-aid drugs in older adults found that for every 13 people who took a sedative, like Ambien, one had improved sleep — about 25 minutes per night on average — while one in six experienced a negative side effect, with the most serious being increased risk for car accidents.

“There’s this cognitive dissonance, or almost professional depression,” Walker says. “You think, ‘Oh my gosh, I’m a doctor, I’m going to give all these drugs because they help people.’ But I’ve almost become more fatalistic, especially in emergency medicine.” If we really wanted to make a big impact on a large number of people, Walker says, “we’d

be doing a lot more diet and exercise and lifestyle stuff. That was by far the hardest thing for me to conceptually appreciate before I really started looking at studies critically.”

Historians of public health know that most of the life-expectancy improvements in the last two centuries stem from innovations in sanitation, food storage, quarantines, and so on. The so-called “First Public Health Revolution” — from 1880 to 1920 — saw the biggest lifespan increase, predating antibiotics or modern surgery.

In the 1990s, the American Cancer Society’s board of directors put out a national challenge to cut cancer rates from a peak in 1990. Encouragingly, deaths in the United States from all types of cancer since then have been falling. Still, American men have a ways to go to return to 1930s levels. Medical innovation has certainly helped; it’s just that public health has more often been the society-wide game changer. Most people just don’t believe it.

In 2014, two researchers at Brigham Young University surveyed Americans and found that typical adults attributed about 80 percent of the increase in life expectancy since the mid-1800s to modern medicine. “The public grossly overestimates how much of our increased life expectancy should be attributed to medical care,” they wrote, “and is largely unaware of the critical role played by public health and improved social conditions determinants.” This perception, they continued, might hinder funding for public health, and it “may also contribute to overfunding the medical sector of the economy and impede efforts to contain health care costs.”

It is a loaded claim. But consider the \$6.3 billion 21st Century Cures Act, which recently passed Congress to widespread acclaim. Who can argue with a law created in part to bolster cancer research? Among others, the heads of the American Academy of Family Physicians and the American Public Health Association. They argue against the new law because it will take \$3.5 billion away from public-health efforts in order to fund research on new medical technology and drugs, including former Vice President Joe Biden’s “cancer moonshot.” The new law takes money from programs — like vaccination and smoking-cessation efforts — that are known to prevent disease and moves it to work that might, eventually, treat disease. The bill will also allow the FDA to approve new uses for drugs based on observational studies or even “summary-level reviews” of data submitted by pharmaceutical companies. Prasad has been a particularly trenchant and public critic, tweeting that “the only people who don’t like the bill are people who study drug approval, safety, and who aren’t paid by Pharma.”

Perhaps that’s social-media hyperbole. Medical research is, by nature, an incremental quest for knowledge; initially exploring avenues that quickly become dead ends are a feature, not a bug, in the process. Hopefully the new law will in fact help speed into existence cures that are effective and long-lived. But one lesson of modern medicine should by now be clear: Ineffective cures can be long-lived, too.



David Epstein covered science and medicine issues as well as sports science. Prior to joining ProPublica, he was a senior writer at Sports Illustrated.

Illustrations by Giulio Bonasera, special to ProPublica. Art direction and production by David Sleight.
