

## Here's what happened when a promising clinical trial for depression was halted

A

n early halt to a trial of deep brain stimulation for depression reveals little about the treatment but more about the changing nature of clinical trials.

Some medical experiments are more daunting than others. The one that neurologist Helen Mayberg came up with to test a model of depression she had developed over about 15 years involved drilling two holes in the top of a patient's skull and sliding two low-voltage electrodes deep into the brain until they reached a region known as Brodmann area 25. In a pair of pale pink curves of neural flesh called the subcallosal cingulate, each about the size and shape of a newborn's crooked finger, area 25 occupies the fingertips. Once in place, the electrodes were wired to a battery pack implanted in the patient's chest and turned on.

Between 2003 and 2006, first at the University of Toronto and then Emory University in Atlanta, Mayberg tried this experimental surgery, called deep brain stimulation (DBS), on 20 people who had been torturously depressed for years despite trying every other possible treatment. Area 25, she believed, was a neural junction box that became hyperactive in depressed people. She designed the experiment to see if the thrum of these DBS devices would calm area 25 and relieve their despair. In 2005 she published results from the first six patients, who'd done amazingly well; the other 14 later showed similar progress. Most cut their depression scores by around half. Over a third left their depression behind entirely.

Mayberg's 2005 paper, published in *Neuron*, amazed her colleagues. To get such results in patients this sick was stunning. Tom Insel, then director of the US National Institute of Mental Health, called it "a new way of understanding depression". The attention expanded into the public realm in 2006 – a story I wrote about Mayberg's work in the *New York Times Magazine* that April; a CBS News *60 Minutes* segment that September. Her phone rang ever more often with calls from reporters who wanted to interview her, universities and institutions that wanted her to come lecture – and scores of hopelessly depressed people, or their family members, who'd seen the press and wanted the treatment. Mayberg was on fire.

It wasn't long before St Jude Medical, the device company that made the neurostimulators Mayberg used, decided to run a clinical trial to seek US Food and Drug Administration (FDA) approval for the treatment. There were, of course, risks to moving ahead so soon. Mayberg and her colleagues, for instance, had been using the implants with more people and were seeing hints of variables that might affect the treatment's power – differences in the quality of the patient's depression, in how the surgeons sited the device – that if explored might make the treatment even more effective. You could of course wait for ever – you could always find out more – but at some point you had to forge ahead. And the thing was *working*. People were all but rising from the dead.

So in 2008, after much discussion, she signed on as one of more than two dozen researchers from all over North America who would serve as consultants and investigators to help design and execute the trial. As is usual in trials of this size, with multiple treatment centres, the sponsor – St Jude, in this case – would actually administer it. Many of the researchers had done trials before; Mayberg had not. And no one researcher, including Mayberg, could hope to direct everything. She hoped to do what she could to ensure the trial was rigorous and the patients treated well.

St Jude branded the trial the Broaden study – for Brodmann Area 25 Deep Brain Neuromodulation. The plan was to have 20 neurosurgery centres implant St Jude's DBS devices in 200 people with treatment-resistant depression, then turn on the devices in two-thirds of them. In the other third – the control group – they'd leave the devices for six months before activating them. No one with patient contact would know which patients were which. Then they'd follow all 200 patients for two years after implantation to see how the two groups fared. Would the active-implant group do better than the control group? Would they do as well as Mayberg's own patients?

By 2017, when it was over – long, long after it should have been – Broaden would reveal not only the answers to those questions, but also deep tensions and confusion about the limits of a clinical trial, the nature of research, and the difference between the two.

It started well, as three of the trial's surgical teams worked steadily in 2008 and early 2009 to implant the first 30 patients. At that point, as planned, there was a mandated safety review. After a pause of some months to generate a report and review it with the FDA, the implantation programme expanded from three surgical centres to 15.

As data from all those new patients implanted at new centres trickled in over the following months, it became apparent that while many were becoming less depressed, they weren't progressing as rapidly as Mayberg's previous, non-trial patients had. This wasn't surprising. Large, double-blind, control-group trials often fall short of the uncontrolled "open-label" work that inspired them. ("Open-label," or unblinded, means all patients know they're getting the treatment.) So these early results didn't much worry Mayberg.

As the 15 surgical teams implanted yet more patients in 2011 and 2012, the study neared the mark that would trigger its first assessment – an interim "futility analysis" that would evaluate progress when at least 75 patients had been implanted for six months or more, with two-thirds in the active-treatment group and one-third in the control. As it happened, a large group of patients reaching the six-month mark at about the same time allowed the team to include 90 patients in this interim analysis. A critical question in this analysis was how the patients' progress measured up against the trial's primary target, or "outcome measure". In an FDA trial of this sort, the primary outcome had to be declared in the study design; it had to be relatively simple and clear, to prevent the moving of goalposts or other hanky-panky; and it had to reflect the therapeutic goal, which in this case was to reduce depression.

[mosaic](#)

least 40 per cent of the active-group patients to experience a drop of 40 per cent or more in their individual depression scores. The scale used was a common, clinician-administered depression questionnaire called the Montgomery–Åsberg Depression Rating Scale. In this 10-question, 60-point scale, practices in categorising severity vary, with typical schemes rating scores under 9 or 10 as indicating no depression, around 10 to 18 mild depression, 19 to 34 moderate depression, and 35 and up severe depression. The Broaden trial, which was interested mainly in change over time, classified scores under 11 as no depression (or remission), and didn't categorise severity otherwise.

The Broaden patients, having tried virtually every approved therapy without lasting benefit, had gone years with scores averaging in the mid-30s. For them, a 40 per cent drop would still leave them depressed, but not nearly so badly – an ambitious goal, given how immovably sick they had long been. Yet this seemed reachable given Mayberg's impressive previous results.

As well as looking at the 40 per cent improvement target, the Broaden team were also keen to see how many patients would achieve full remission; would they match the rates seen in Mayberg's open-label work?

None of it happened. In late 2012, when the team finally had six months of data on 90 patients, the depression score reductions weren't even close. Of the 60 active patients, only 20 per cent (12 of them) had reduced their scores the desired 40 per cent after six months, and only 5 per cent (three people) were in remission. This was essentially matched by the control group, who after six months with inactive devices had 17 per cent hitting the improvement target and 7 per cent in remission.

While discouraging, these numbers did not automatically sink the trial. Rather, this interim analysis was done to gauge the trial's likelihood of hitting its outcome goals if it implanted 200 patients for at least half a year. But if the calculated likelihood fails to clear a trial-specific, predetermined minimum set by the FDA – a sort of kill switch – the FDA will usually end the trial with a so-called “halt for futility”.

For the Broaden trial, the FDA had set this bar at a likelihood of success of one in ten. The futility analysis calculated its likelihood of success as one in six. Modest odds, to be sure, but it easily cleared the FDA bar. The trial could continue.

However.

A trial sponsor can stop a trial at its own prerogative. St Jude now did so. Sometime in 2013, the company informed the Broaden team – Mayberg and dozens of cooperating researchers at the 15 research centres – that it was stopping the trial. St Jude never said why. DBS procedures can cost upwards of [\\$100,000 a patient](#), so continuing would have been expensive for St Jude. Perhaps they had their own kill-point somewhere north of one in six. In any case, they stopped the trial.

Image not found or type unknown

Image credit: Eleanor Shakespeare for Mosaic

St Jude and then Abbott, a larger company that later bought St Jude, would continue to pay for the care of the implanted patients who wanted to either continue or have their devices removed. They'd pay the research centres to continue to collect data until every patient had hit the two-year mark. But they would implant no more patients, and they would cease their attempt to get FDA approval. Five years in, roughly ten million dollars out, and a decade after Mayberg implanted her first horribly sick patient, the trial was over.

The first reaction to this halt – to the end of this keenly anticipated test of a depression treatment hailed as the most original and promising in decades – was crickets. St Jude said nada. The researchers, who'd agreed not to talk until the data was published, zipped it. No one associated with the study publicly marked its halt.

Yet word slowly slipped out.

The first leak came via industry analyst James Cavanaugh, who in late December 2013 reported in his newsletter, *Neurotech Business Report*, that he had learned at a scientific meeting that the Broaden trial had “failed a futility analysis”. This news, Cavanaugh wrote, “cast a pall over an otherwise upbeat attendance... Once again, the industry is left to pick up the pieces as a promising new technology gets set back by what could be many years.”

The rest of Cavanaugh's post was nuanced and carefully considered. But the language used by his unnamed source – “failed a futility analysis” – strongly implied that the trial was halted by the FDA because it did not

clear the FDA's chance-of-success threshold. It didn't help that "failed a futility analysis" just sounds disastrous because in common parlance, if not in the FDA's, "futile" means "pointless". Cavuoto made clear in his post that he thought Mayberg's treatment was useful and promising. But in the silence left by the company and its shushed researchers, his "futility" phrasing set the tone for much of the coverage and conversation that followed.

A handful of journalists and bloggers who had followed the trial soon waded in, some more carefully than others.

In January 2014, a month after Cavuoto's post, a blogger known as Neurocritic, one of the sharper online writers on neuropsychiatry and a longtime follower of Mayberg's work, [reported and amplified Cavuoto's scoop](#). Neurocritic's measured, context-rich post inspired a measured, context-rich stream of comments that accumulated beneath the post over the next two-and-a-half years. These comments included numerous remarks from about 10 to 12 people identifying themselves as patients in the trial. (An exact count is impossible, as many are signed "Anonymous" and some appear to be follow-up comments.) Two, three, or possibly four said the treatment didn't help, and one or two reported severe complications of the sort warned of in any trial involving such surgery but that are nonetheless withering to experience and distressing to read about. Somewhere from seven to nine commenters said the treatment had worked for them or their spouses; several called it life-saving.

In March 2014, the *Scientific American* online columnist John Horgan, who had previously written critically of Mayberg's work (and – full disclosure – of some of my coverage of it), posted a brief report on the halt that replicated Cavuoto's phrasing about "failing the futility analysis". Two readers commented that they'd been in the trial and it had saved them; a third wrote saying his wife had been in the study and suffered "severe memory and cognitive problems as a result". (These people may or may not have been among those who commented at Neurocritic's site.) Horgan's take was factually accurate, if sharply critical. Others drew on mischaracterised or mistaken reads of the sparse data to paint the trial as a massive failure. Speculation filled the data void.

Mayberg, reading such coverage and comments and hearing of rumours flying around among colleagues, was appalled. "I was mortified. I asked the company, *Please* make a public statement, because all you have now is a rumour mill and the information that a few patients weren't doing well." St Jude stayed mum. Before long, colleagues at conferences were offering Mayberg condolences. They'd pull her aside and say they were so, so sorry to hear the trial had failed.

One day she got an email response to a grant application in which the grant officers asked *Why are you still working on this if it doesn't work?* Others said similar things. *Why write about this*, a journal editor asked in a letter rejecting a paper she'd sent, *when everybody knows it failed?* She began to worry her entire research programme was endangered.

Then something emerged that both sharpened and salved the sting of the trial's halt. The patients got better.

To be sure, the treatment didn't work as well as hoped. As expected, it failed to help many. In the first

year, ten of the 90 patients left the study (and four had their devices removed), for reasons ranging from worsening depression to a suicide attempt. Eventually, of the 90 patients, 37 – most of those who'd felt no benefit – had the devices removed.

Also as expected for a surgical intervention in so depressed a population, some experienced side-effects and complications of the sort mentioned by commenters on the Neurocritic post. Overall, at least nine people reported increased depression, six got infections, and several more suffered side-effects such as headaches or post-operative discomfort or pain, either at one of the surgery sites or from a condition called “bowstringing”, in which the battery leads running under the skin between chest and skull bond with tissue and create pulling sensations. Three patients grew more anxious. One considered suicide. One tried it and lived. Two tried it and died.

Perhaps the best known of those with complaints was Steve Ogburn, an architect who was implanted at Stanford in late November 2012 and soon experienced complications. In a video he posted in 2014, and cited by John Horgan and several others who then wrote about him, Ogburn reports a harrowing tale of “severe cognitive decline”, continued depression, and severe pain in his head and around the electrical leads running between chest and head. None of these, he says, were relieved when Stanford removed the implants, wiring and battery in December 2013.

Lisa Wick, an elementary school teacher from Minnesota, also suffered complications, but hers developed seven years after what had otherwise been a successful implant. When she was first implanted, in 2008, Wick had been badly depressed for years despite psychotherapy, several antidepressant regimes, and multiple rounds of electroconvulsive therapy. “*Just* able to survive,” she told me. “I’d teach, then go home and crash.”

Like many other patients who’ve improved with DBS, Wick vividly remembers feeling a sudden change right on the operating table, which she feels certain was the stimulator being turned on. “All of a sudden I was talking. It was like that part in *The Wizard of Oz* where it goes from black-and-white to colour. I felt better right away.” The pain she’d felt for ages quickly lifted.

Within a year, Wick’s depression eased to a level she was happy with, and she stayed well until an infection that developed after a battery change in 2015 spread up the leads towards her skull and eventually forced the Chicago team to remove the entire implant.

With the implant gone, Wick slid back into the paralysing depression and unceasing psychic pain she’d felt before the surgery. After almost a year of this, with Mayberg’s help, she was reimplanted by a Broaden team in Dallas. According to Mayberg, that team was able to duplicate exactly the first implant’s location and settings. Wick was unconscious this time around, but when she woke, she found the “awful, awful” active pain that was crushing her had vanished. Over the next year she climbed back into the state of full remission she’d had before the infection. “A long wait,” she told me in December 2017. “But I’ve returned to the place where I longed to be. I’m back.”

Stories like Wick’s and Ogburn’s, though mere anecdotes from a statistical perspective, are hardly meaningless. They give a taste of how various, variegated and rich these cases can be, how fickle is fate

– and biology. They speak to the tremendous stakes involved in aiming a treatment so intrusive at an affliction so crushing. Yet in the brutally binary terms dictated by a clinical trial, the significance of such cases is largely limited to how well they represent the study’s group-level data. And under those terms, each story represents but one point of data among 90; Wick and Ogburn were dots near the opposite ends of a spectrum. And going by the six-month data, as Mayberg herself is quick to confirm, “the trial clearly failed”.

But what about the treatment? In the months and years after the November 2013 halt, as data accrued from patients who continued with the treatment, it became clear that more and more of them were moving towards and past the 40 per cent improvement threshold; some were even in remission.

Most of the researchers had access to reports about the data as it accrued. But the first presentation of results – a presentation sharply constrained by journal guidelines, St Jude’s proprietary hold on the raw data, and authorial arguments about relevance – was made public only in October 2017.

The paper, finally published in [\*Lancet Psychiatry\*](#), presented patient data through 24 months of activation. When tracked for two years instead of the six months used for the futility analysis, the percentage of active-treatment patients whose depression scores dropped by at least 40 per cent more than doubled, to 50 per cent of all those in the original active group. The remission rate also rose, from 10 per cent at six months to 31 per cent at 24 months. These rates come very close to Mayberg’s open-label trials.

Also encouraging was the response of patients in the control group when their implants were activated after six months, for they continued to roughly match that of the group whose implants had been active from the start. The control group’s full-remission rate once active, however, was lower. This may be because the control group’s smaller size can magnify the statistical impact of small differences in the depression scores of those patients. We don’t know, for instance, whether a high number of control-patient scores fell just short of remission (or, conversely, barely reached it) – just one of many things we can’t know because neither St Jude nor Abbott has published the individual-patient data.

In any case, this long-term Broaden data echoes Mayberg’s findings, in several open-label studies done before and during the time of the Broaden trial, that the treatment gains effectiveness over time. The two-year results for these intensely sick people – half reaching the 40 per cent improvement threshold, almost a third in remission – stand sharply at odds with the six-month scores.



Image not found or type unknown

Image credit: Eleanor Shakespeare for Mosaic

What we have here is a failed clinical trial – of a treatment that seems to work.

It's a nasty conundrum. As Paul Holtzheimer, a lead author on the Broaden study, put it at a conference last year, "To imagine that 50 per cent of patients that are this severely ill, this treatment-resistant, would get better and stay better for this period of time... [We] have [a large failed study], we have these really amazing open-label pilot data – it is hard to reconcile those two."

Holtzheimer's lament touches on two separate but connected problems. One is why the trial failed even though the patients got better. The other is why it fell short of Mayberg's open-label studies.

The simpler question first: Why did the trial fail even though the patients markedly improved?

Perhaps the trial's sharpest constraint was the weight placed on its six-month results. Although six months is actually long for a depression trial (trials of pharmaceutical antidepressants, for instance, typically last only six to 12 weeks), it makes a pretty short leash for Mayberg's DBS approach, which earlier work had shown led to marked improvements between six and 24 months. A study of her first cohort of patients, operated on in Toronto, found the percentage who markedly improved – that is, cut their depression scores by at least 40 per cent – was 62.5 per cent after one year, 46.2 per cent after two years, and 75 per cent at three years; remission was achieved by 18.8 per cent at one year, 15.4 per cent at two years, and 50 per cent after three years. A study of another 17 open-label patients Mayberg treated at Emory found improvement in 41 per cent of patients at six months, which rose to 92 per cent at two years; remission rose from 18 per cent at six months to 58 per cent at two years. Betting the entire Broaden trial



on favourable six-month results proved far riskier than anticipated.

So why just six months? According to several Broaden study authors, the six-month mark was set because of concerns that, given the treatment's high effectiveness in previous studies, it would be unethical to keep a third of these very sick patients on placebo for any longer. A shorter stretch would also save St Jude a lot of money if the trial did need to be stopped. In any case, Broaden's six-month outcome target forced an early, one-time-snapshot evaluation of a therapy that did its best work over longer periods. It also meant losing the control group from any analysis beyond six months, making it difficult to assess later progress objectively.

The second question is more involved: Why do patients in Mayberg's open-label work – especially those she has treated since Broaden started – have such better response rates than those in the trial?

One possible contributor is bias. Researcher bias, which can take many forms, is a potential force in any study. It's possible – arguably somewhere between likely and certain – that Mayberg's strong role in her open-label work affected the outcomes. As is standard practice, the depression scoring was done by researchers with no ties to her or her lab. Yet Mayberg is a warm, confident, highly engaged clinician, and the bonds she builds with her patients doubtless give many of them more confidence and faith in the procedure.

As some observers have noted, Mayberg had two conflicts of interest that could bias her work: licensing and consulting fees from St Jude, and part ownership of a patent on the procedure that could generate income should it get approved. During the trial, she and Andres Lozano, the Toronto neurosurgeon who shares the patent with her, received fees from Abbott for a licence to use the procedure in the trial. Mayberg says her own licensing and consulting fees during the trial amounted each year to “about enough to buy a decent used car”. Decent-used-car money is not meaningless. But it is modest next to the shiny-new-Lexus-level fees common in clinical trial consulting agreements, and it falls far short of what Mayberg could likely demand and receive.

Follow the latest news and policy debates on sustainable agriculture, biomedicine, and other ‘disruptive’ innovations. Subscribe to our newsletter.

[SIGN UP](#)

A more obvious difference was that the open-label work had no control groups and the Broaden trial did. As every Broaden patient was told beforehand, one in three of them would start the trial with a device that would be inactive for the first six months. Possibly this knowledge created a sort of bias within them – a reduced confidence and thus response – in both the placebo and the active groups.

The factors above may well have contributed to the different responses in the trial and the open-label work. But it seems a stretch to think they could make Mayberg's open-label work twice as effective over six months as the trial was. The answer is more likely found in the many differences between Mayberg's open-label treatment protocol — which changed as she went along — and the protocol in the Broaden trial.

Even as the Broaden trial was running, Mayberg continued her research, implanting new patients and, in

an effort to get ever-better results, refining the way she was using DBS. This meant that her work increasingly diverged from the Broaden study in significant ways. She selected patients differently. She sited and managed the electrodes differently. She supported the patients differently. Any of these had the potential to create or widen a gap in outcome between Mayberg's open-label work and the Broaden trial. In a sense, that was the point: She wanted to improve on the early protocol on which Broaden was largely based.

The trial and the open-label work differed in how they chose patients. To start with, the Broaden patients had suffered from treatment-resistant depression far longer than Mayberg's patients had – on average, 12 years instead of five. This longer run may have contributed to their reduced response.

And while Broaden filtered out people with anxiety or personality disorders (as did Mayberg), it did not exclude or prioritise any depressive subtypes. Meanwhile, Mayberg had begun to identify in her open-label patients at least three psychological characteristics that predicted better response: comparatively low "mood reactivity", or short-term responses to changing environmental conditions; a clear "psychomotor" slowing of thought and movement; and "high negative affect", or the experience of depression not just as an absence of pleasure but as the distinct mental or even physical pain that the psychologist and philosopher William James, who often suffered depression, called "an active anguish" – the pain that Lisa Wick felt lift both times her device was implanted. These traits characterise what Mayberg calls a "classically melancholic" depression, in which calming area 25 may have more effect. Mayberg now selects for such patients; the Broaden study neither sought nor avoided them.

Another difference was that the Broaden study sited the electrodes by a method that Mayberg happened to abandon as the Broaden study began. Until 2008, Mayberg's team sited the implants based on so-called gross anatomy. That is, they used conventional MRI scans to place the electrodes in a particular spot within the well-defined brain area, visible in any scan, called area 25. Broaden targeted its placements likewise.

Yet even as Broaden launched, Mayberg began using a newer, more detailed imaging tool, diffusion tensor imaging (DTI), that could reveal not just distinct brain areas but the white-matter bundles, called tracts, that carry neural traffic from one area to another. This [DTI work showed](#) that patients responded best if the electrodes sat at particular junctions of white-matter tracts within or next to area 25. Together with adjustments in patient selection, this has helped to improve outcomes in Mayberg's open-label patients since 2008 – but came too late for use in the Broaden study.

Finally, Mayberg has long offered her patients comprehensive, personally tailored programmes of psychiatric and social-service support to help them rebuild their lives. After years of deep depression, most patients' lives, relationships and ways of thinking have been entrenched in illness and disability long and deeply enough that surgery alone was unlikely to make them whole. Like a patient with a knee repair, says Mayberg, "they need rehab to get well again". So her team helps them get psychotherapy, occupational or physical therapy to rebuild skills or physical health, and other assistance to connect them to needed social services.

The Broaden trial offered only post-surgical support, and to prevent statistical confounds it expressly ruled

out the addition of any psycho- or drug therapy not underway before the trial. This successfully isolated the surgical intervention statistically – but likely reduced the chance of recovery.

To look at the Broaden trial in light of Mayberg's larger hypothesis and work is to see that Broaden was a pale outline of the treatment it set out to test – recognisable in its main features, but not to be mistaken for the real thing. Running a clinical trial, of course, all but requires that we strip an experimental therapeutic programme down to its most basic, reproducible and scalable parts. But it would be foolish to take the results of that necessarily reductive exercise as a final verdict on all possible versions of the treatment. A trial of aspirin for headache, for instance, could fail if the patients had migraines, took doses too small, or were assessed for effect after 5 minutes instead of 5 hours.

Mayberg devised her treatment to test her hypothesis, developed after years of scanning work, that area 25 is a crucial node (and thus a possible treatment target) in a neural network critical to depression. The encompassing hypothesis – that such networks play a crucial role in at least some mental-health disorders – suggests that other nodes, treated either alone or in combination with area 25, might also be good targets. Neither of these hypotheses comes close to being proven wrong (or right) by the Broaden trial, any more than a failed aspirin trial would disprove the notion that a headache is at least partly a neurochemical event.

These limits to clinical trials' implications apply tenfold to any trial halted early. Maureen Meade, a critical care specialist and researcher in clinical trial and research methods, put the case against stopping trials for futility in 2005. She wrote that it impoverishes data not only about the primary outcome measure, but also about secondary measures, adverse effects, or subgroup effects that could be informative scientifically even if the primary target is not reached. Halts for futility thus "increase the risk of misinterpretation", including the false conclusion that an effective treatment is not.

The whole idea of "futility" is to stop trials that seem almost certain to fail. Meade's article and the wider halt-for-futility literature make clear that such a halt should be a scientific decision, not a business decision. Thus the usual method described is an interim analysis by a committee of scientists to see if the trial hit a predetermined early benchmark – exactly the kind of review that Broaden passed. Yet St Jude halted it, not the researchers; and at least publicly it cited not any benchmark, nor any data, nor anything at all. As Meade put it, halt decisions made without following clear standards make it more likely that "stopping decisions will be idiosyncratic or self-interested". Whether St Jude's halt was idiosyncratic or self-interested can't be said. But the decision can hardly be considered scientific, because science means showing your data. Today, many companies involved in clinical trials are being more open with data from their studies, even those that 'fail', recognising the value of letting other researchers use it to build greater understanding and make more progress faster.

Cast in terms of a courtroom trial, then, the halt of almost any clinical trial – and clearly this one – declares not a verdict but a mistrial. The question at hand remains unresolved, to be answered another day. Both Mayberg's treatment and her wider hypothesis must still prove themselves. But the Broaden trial has hardly proven them wrong.

Will Broaden's failure put an end to talk of DBS for depression? Will it jeopardise Mayberg's research

programme?

This no longer seems likely. After a period in which Mayberg had reason to worry her funding might evaporate, the creation of the National Institutes of Health's Brain Initiative in 2013 led to a friendlier funding atmosphere for her work. In 2017 the National Institute of Mental Health awarded her a \$5 million-plus grant to continue her DBS work. And in January 2018, she moved from Emory to the Medical School at Mount Sinai Hospital in New York, where she heads the new Center for Advanced Circuit Therapeutics. There she continues to refine both her hypothesis and her treatment.

Whether any device maker will think it worthwhile to fund another clinical trial of area 25 DBS for depression remains to be seen. Abbott still holds the licence to treat depression in this way. James Cavuoto told me he thinks Abbott likely remains interested. Abbott told me it has "ongoing dialogue" with researchers in DBS and depression.

If Abbott decides to take another shot at area 25, it will be able to work with more flexible trial rules created by the FDA in 2016. These rules allow a study to designate several different outcome measures, including effectiveness measures on different timescales rather than at a single time, to add additional outcome measures based on new hypotheses (which might dictate a change in target or differences in device placement), and to identify and evaluate the characteristics of any subsets of the trial population that are responding particularly well.

Clinical-trial watchdogs rightly note that using too many measures or outcome targets or patient subsets can open the door to cherry-picking or moving the goalposts. But if controlled well, such flexibility could reveal whether a treatment or a variation of a treatment might work well for one population but not another – in people with a particular depression subtype, for instance. This would be roughly akin to the approval of cancer drugs that work on, say, only 1 in 5 people with a broadly defined type of tumour, but in 4 of 5 people whose tumours include a specific genetic signature.

Taken together and managed well, the new rules should allow more flexibility in trials of complex treatments like Mayberg's. In a way, these rules seek to find a better way to balance what machine-learning people call the explore–exploit trade-off. This is the tension between exploring all possibilities in some arena and finding the quickest way to exploit it. A trial like Broaden is forced to focus on a single point of exploitation. A trial with broader parameters could test several different outcome measures or scales instead, and thus bring at least a bit of the exploratory nature of research into the game.

Cavuoto hopes Abbott will run another trial and try to answer those questions. The FDA rule changes, the continued improvement of the Broaden patients, and the potential gains, he says, make it too ripe to pass up. From a business point of view, he notes there's "a huge market opportunity. But more important, there are thousands, perhaps millions of patients who are suffering that might not be. This disease is literally killing people."

In the meantime, says Mayberg, “I just need to keep working.” She wants to identify the success factors – variables in patient phenotype, electrode placement, post-surgical support – well enough that they’re refined into a more transferable protocol.

“My husband asks me when I’m going to retire. I’ll retire when the work is done.

“I can’t stop because some people think it’s worthless. My patients tell me these people are wrong. The data tells me they are wrong... This thing isn’t dead.”

**David Dobbs writes about science, medicine, sports, fishing, music, human behaviour and other forms of culture. He has written for [publications](#) such as National Geographic, the New York Times Magazine, The Atlantic and Pacific Standard. Follow him on Twitter [@David\\_Dobbs](#).**

**A version of this article was originally published on Mosaic’s website as “[What can we learn when a clinical trial is stopped?](#)” and has been republished here with permission.**