THE PRICE WE PAY

WHAT BROKE AMERICAN HEALTH CARE — AND HOW TO FIX IT

MARTY MAKARY, MD

The *New York Times* bestselling author of *UNACCOUNTABLE*

BLOOMSBURY

CHAPTER 7

Dear Doctor

The usual odorless, tasteless cubed cantaloupe made its standard appearance in the back of the Marriott Marquis meeting room, but this was no typical meeting. I was in Washington, D.C., with the top leadership of an association of skin cancer surgeons called the American College of Mohs Surgery. If you're not one of the millions of patients a year who get skin cancer, you may not have heard of Mohs surgery. But the technique—developed in the 1930s by Dr. Frederic E. Mohs[1]—is a big reason skin cancer is much more manageable today.

I had heard that the Mohs surgeons were interested in addressing overtreatment. And months earlier, I'd connected with the association's president, Dr. John

Albertini, by phone. I had told him how my Hopkins team and I wanted to work with specialty associations to identify ways to measure the appropriateness of medical care. Right away, he got it. He even one-upped me, telling me about something else he saw doctors doing that troubled him.

To enable me to understand the problem, Albertini had to explain the technicalities of Mohs surgery. In this specialty, the doctor's role is unique because the doctor acts as surgeon, pathologist, and reconstructive specialist all during the same procedure. The goal is to excise all the skin cancer while minimizing the amount of healthy flesh that gets removed. A Mohs surgeon starts by cutting out the cancer in a block of tissue and examining it under the microscope. It might be a sliver of flesh or an inch-square cube. If the tissue block has cancer cells on the edge, that means the surgeon didn't remove all the cancer. It's what we surgeons call a "positive margin." The surgeon goes back to the patient and removes an additional sliver of tissue at that location. Each tissue block removed is referred to as a "stage" of the surgery. The Mohs surgery breakthrough is re-

moving all the visible cancer while preserving as much normal skin as possible. In the old days, skin cancer surgery disfigured patients because doctors took so much flesh at once.

Here's where things get interesting. Mohs surgery typically takes one or two precise stages. On rare occasions, a third stage may be necessary. Surgeons get paid well for Mohs procedures. And it turns out that they get paid per stage. Cut a little extra here and there and you get a bigger paycheck, whether the extra cuts are necessary or not. As a surgeon, I was familiar with these types of financial "carrots" lying around the operating room.

Albertini explained that over the last several years, the association's leadership had heard multiple reports that some doctors appear to be doing the operation in too many stages. It may be that the doctors need further training. Or they could be motivated by money. Albertini proposed a pattern we could examine for the Improving Wisely project. We could look at the average number of stages each surgeon used during the procedures. We would see who was making the most cuts.

“Most surgeons fall in a certain range,” Albertini told me. “But some are going to be way out there adding time and expense to the procedure and unnecessary surgery for patients.”

Albertini said that the project’s success would depend on the buy-in of his colleagues: the leaders of the Mohs surgery association. We made a plan for me to join the association’s executive leadership team when they gathered at the American Dermatology Association conference in Washington, D.C.

A Crucial Buy-in

I was nervous walking into the hotel conference room, so I took a pass on the cantaloupe cubes. This was my first time pitching to a surgical society the idea of analyzing the practice patterns of individual doctors. I hoped they would agree that lowering health care costs can start with eliminating medical care that doesn’t need to be done in the first place.

This meeting could be critical. We as a country spend more than $15 billion a year reporting quality metrics

to the government and to one another.[2] Doctors tire of flavor-of-the-month quality improvement campaigns —especially those imposed on us without our input. I knew doctors on the front lines had to define which practice patterns were appropriate and which were not. I needed to see consensus from them. I needed them to tell me the best way to proceed.

Albertini welcomed me, introducing me to the titans of the field. Sitting around the conference table were Dr. Tom Stasko from Oklahoma, Dr. Allison Vidimos from Cleveland Clinic, Dr. Richard Bennett from UCLA, Dr. Victor Marks from Geisinger, Dr. Barry Leshin from Winston-Salem, and Dr. Brett Coldiron from Cincinnati. I had read about Dr. Coldiron: he alone had performed more than 50,000 of these state-of-the art Mohs operations.

I dived into my presentation. I told them about the Improving Wisely model. I explained that the first step would be to identify something that's overdone in their field, then devise a smart way to measure how often a doctor does it, and then see if there is an agreement about how much is too much. Finally, we would reach

out to the doctors whose practice patterns fell outside the boundary of what they considered appropriate. That feedback would show them how they compare to their peers and allow us the opportunity to help these doctors improve. My research team had taken Albertini's idea and run with it. I provided preliminary numbers showing that most surgeons averaged one or two stages during Mohs operations. But some averaged three or four.

As I talked, the surgeons were murmuring and nodding. The vibes were good. Then they started jumping in with comments supporting what I was saying.

"Makes sense," said one of the board members. "We need to do something about surgeons out there who are operating with no accountability."

"There are practice patterns that clearly cross a threshold," another surgeon added.

"This is what a professional association is supposed to be doing," chimed in one of the other Mohs leaders.

They got it! They felt pride in their profession and a duty to act. The surgeons there were concerned that a small number of doctors in their field might be sucking

a lot of money from the system. They agreed that doctors who were out of line would not like being identified as outliers in their field. They hypothesized that these doctors' competitive nature would kick in and they would probably reduce their overuse on their own. From my own observations in medicine, I couldn't have agreed more. People respond well to competition.

The Mohs surgery leaders liked Albertini's proposal, and I did, too. Our intention wasn't to penalize or even require preauthorization for a doctor to remove a cancer in three stages or more; in some cases, it might be necessary. But a pattern of doing three blocks in a large number of patients is something these experts said seemed inappropriate. Cutting through the tumor rather than around it was a lucrative temptation doctors face frequently. Skin cancer is the most common cancer in the world, and the technique is used to treat the basal cell and squamous cell subtypes and is increasingly being applied to melanoma as well.

The Improving Wisely approach sure beat the old way of measuring infection rates and readmissions to the hospital. Both are exceedingly rare with Mohs surgery,

which is done as an outpatient procedure. The board enthusiastically accepted the offer to partner with me and my team at Johns Hopkins. We were ready for takeoff.

Identifying Outliers

I got to work with my research team. We obtained data from the federal government for every patient in Medicare, the government's insurance program for the disabled and patients over age 65. The data included each doctor's identification number, and it showed the number of stages billed for each operation. We used the data to graph each surgeon in the United States by the average number of tissue blocks they removed during the skin cancer operations performed. Sure enough, as the Mohs surgery leaders had predicted, most doctors were within a range of normal practice variation. The typical surgeon averaged between 1.2 and 2 blocks, or stages, per patient over the course of a year. But there were also some outliers who averaged 4 or more stages per patient.

We took our results to the Mohs surgery leaders and they said the analysis confirmed their suspicions. They even recognized the names of some of the doctors who were in the top 2% of the outliers. They had heard stories about them or seen some of their patients for follow-up care. The experts in the field said that any high-volume surgeon who averaged more than 2.2 stages per operation was beyond the threshold of what they would consider appropriate. We had consensus.

Next, we sent letters to about half of the surgeons we analyzed. We didn't reach out to all of them right away because we wanted to study whether our outreach had any effect on their performance. We needed one group with whom we intervened and another that we did not (a control group). The letters came from the American College of Mohs Surgery (ACMS) and my team at Johns Hopkins. They included a one-page report that showed each surgeon how he or she compared to the rest of the Mohs surgeons in the country. A graphic designer helped my team generate physician-specific reports showing where each doctor stood on the bell curve. Doctors who used an average of 3 and 4 stages or more

per case were way out on the tail end of the chart.

We didn't chastise anyone who fell into the outlier category. We simply said "This is where you stand relative to the rest of the Mohs surgeons in the country." We also indicated in the report that the association embraced a range of normal variation in the average number of stages per operation. According to the Mohs experts who designed the reports with us, it would be obvious to the outliers that they were well outside the range of what seemed appropriate.

The letter, signed by the top leaders in the field, also offered educational resources and invited feedback on the project.

We sent out about a thousand reports and then I held my breath. Would the notification make any difference?

Surprising Response

In the days after we notified the Mohs surgeons, I kept expecting the phone to ring with complaints. But the gripes didn't come—neither to me nor to the Mohs surgery leaders who had cosigned the report's cover

letter. Then the emails began to roll in:

- *Thank you for the recent report. I had no idea where I stood relative to my peers nationally and now I know. I'm above average but will take a careful look to see how I can improve.*
- *I love showing this metric to my patients.*
- *I just wanted to give you a quick word of feedback on the Individual Surgeon Data Report I just received: I absolutely love it! I have wanted to know for some time where I stand relative to my peers regarding my average number of stages versus my peers and to my chagrin it just arrived in the mail! It gives us a nice benchmark to how we are doing.*
- *Thanks for sharing this data. I'll work on my technique. Will this information be used for anything? Will it be made public? Please let me know.*

- *I'd like to learn more about the retraining offered by ACMS.*
- *When will the next report be delivered?*
- *Thank you for the report. Very important.*
- *I had heard of the reports coming out and was glad to see I'm not an outlier.*

The surgeons appreciated seeing their data! Sure, the emails were anecdotal, but none of the responses challenged the metric. My team followed up with a survey that found that 80% of all surgeons in the association believed that sharing performance data like this was important. In my opinion, the positive response to the "Dear Doctor" letters was because this program was 100% homegrown, based on the wisdom of practicing doctors who understood the proper use and misuse of their craft.

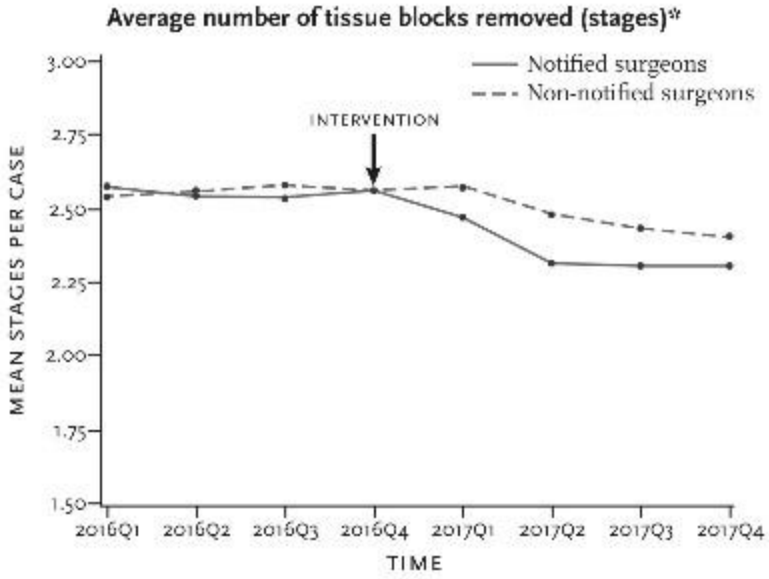
But the million-dollar question was whether our intervention would work. Would it spur outliers to change their practice patterns? Our goal wasn't merely to inform the surgeons; we wanted change. We gave it a

year and checked the national Medicare data again for the doctors we notified. The results were striking. We found that 83% of notified outliers had changed their ways for the better. Moreover, the reduction in blocks per case appeared to be sustained.[3]

The long-term follow-up data revealed an additional interesting trend. In the months after we sent the letters, even the outlier surgeons in the control group, whom we did not contact, began changing their behavior. They had not even seen their data but began to reduce their average number of stages per case, albeit to a lesser extent. Our intervention appeared to have had a crossover effect. And I can see why. It created a lot of buzz when we sent out the reports, and word travels fast among doctors within a specialty. The Mohs leaders also wrote commentaries about the initiative and gave talks about the importance of the program. I heard that some surgeons who fared well in our analysis were broadcasting the fact to their friends and peers. Hey, nothing wrong with that. The program had sent a message to outliers: your national leaders are monitoring the macro trends in your practice data.

Albertini liked what he was seeing. The initiative had created a culture of accountability, and he was hearing stories confirming the improvements were real. "Moreover, no one got humiliated or penalized," Albertini said. "It's a confidential peer-to-peer way to address our outliers in a civil way."

The entire program cost $150,000 that first year, but it resulted in $11.1 million in direct savings to Medicare—that is, to U.S. taxpayers. At the time I submitted this book to the publisher, the savings had escalated to $18 million for the 18 months after the intervention. Not only were the findings well-received by the medical community, but the publication of the results was accompanied by a very supportive editorial entitled "Physicians Respond to Accurate, Actionable Data on Their Performance," written by the American Medical Association board chair Dr. Jack Resneck and University of Iowa Mohs surgeon Dr. Marta Van Beek. The national conversation that followed affirmed the notion that when doctors are involved in quality improvement from the get-go, the results can be incredible. What else in health care yields a 7,430% return on investment?



But why do some performance improvement programs work so well while others struggle? I attribute part of the success to civility. By including practicing doctors early and by using a peer-to-peer method of sharing data in a way that is nonpunitive and confidential, we were highly effective. Moreover, the project focused on what doctors believe to be important. I have visited hundreds of U.S. hospitals and a consistent message I hear from their quality improvement leaders is "We collect all this data, now what do we do with it?" They're burdened with tracking all sorts of things,

some of which don't matter. Ultimately, out of a sense of helplessness, the leaders dump this data on the doctors, who in turn explain it away with the claim "My patients are sicker."

The "My patients are sicker" argument has been a major barrier to improving health care. But it's code for something else. This is doctors saying "You don't understand me or what I do." It's what happens when quality improvement programs are forced on doctors without their consensus. To be effective, a method of measuring care must be developed and endorsed by the doctors and clinicians who work in that specialty. The input needs to come from a range of physicians who serve diverse patient populations. As we expanded Improving Wisely, I required that all the doctors on the expert panels spend at least 70% of their time in patient care. And I insisted on rural and community hospital representation to balance the doctors representing big academic hospitals.

Many other industries have their practice patterns measured. In 2009, the utility company Positive Energy (now Opower) was interested in reducing power use in

neighborhoods. Their data showed that some households used far more electricity than their neighbors. After all, there are no standardized protocols on turning lights on or off when one vacates a room. Just ask anyone who's argued with a spouse about this issue.

The company decided to mail each household a regular feedback report that compared their electricity and natural gas usage to that of similarly sized households in their neighborhood. Playing on the benchmarking theme, the data feedback intervention resulted in an overall reduction in household energy use. When people saw they were outliers, they modified their habits so their usage fell more into line with that of their peers. In a year, this simple intervention reduced the total carbon emissions of the participating houses by the equivalent of 14.3 million gallons of gasoline, saving consumers more than $20 million.[4] Lots of utility companies now take this approach—and it works.

Metrics Matter

I've examined hundreds of quality metrics over the

years and developed my own. I've come to believe many of them need context to be meaningful. The metrics must zero in on what it means for a patient's quality of life and potential disability. The criteria should focus on significant harm or waste by extreme outliers rather than small variations in practice. The metric also must be measurable and designed so it can't be tainted by bias or gaming. And finally, a sound metric should be highly actionable for the physician. Metrics such as mortality, while easy to collect, are hard to make actionable. We need more measures that provide direct insight into what the individual physicians can do right now to modify the way they practice.

After the skin cancer project with the American College of Mohs Surgery, I asked its leadership if anyone had ever proposed the measure they used for the project to Medicare or the broader health care community. Our data suggested that reining in these unwarranted practice variations has dramatic implications for lowering health care costs. "No," they said. No one in the broader medical policy world ever asked for their input, they explained. Yet again, I saw the discon-

nect between those making the rules in health care and those practicing it.

To try to address this gap, I set up a meeting with midlevel Medicare leaders. They liked the idea of setting boundaries of acceptable variation. They referred me to their website, which lays out their standard process for proposing a new quality measure for Medicare to consider using. I noticed that one of the requirements was that any proposed new measure be supported with multiple published articles proving that the measure was evidence-based. That's a nice idea, but a narrow way to look at quality improvement. No one would ever do a trial comparing the outcomes of surgeons who take two versus three blocks per case. For one thing, a trial subjecting people to three blocks per case would be unethical. I gave up on the website.

Medicare requires a quality measure to be based on published evidence. But in the Robert Wood Johnson Foundation project on Mohs surgery, we had a different way of looking at things. We used the wisdom of busy practicing doctors to create a specialty-specific way to measure quality.

The following year, I was invited to meet with the new batch of Medicare leaders at the highest level, Seema Verma, Paul Mango, Kim Brandt, and Adam Boehler. I explained to them that these pattern measures were very telling and that using them had broad implications for cutting waste in Medicare. Waste like the unnecessary vascular procedures Medicare was paying for, such as the ones being performed on church members down the street. The Medicare leaders I met with quickly "got it" and made further work on pattern measurement a priority. Within months, Medicare made plans to send out "Dear Doctor" letters to the country's most extreme outliers.

The peer comparison program conducted by the American College of Mohs Surgery was extremely well received. The following year, the association's leadership decided to expand the program to tackle overuse of skin flaps (a technique to move skin) and the overuse of Mohs surgery in areas of the body where it was rarely indicated, such as lesions on the trunk and legs.

The Improving Wisely project looking at skin cancer surgery rejected the conventional way that we measure

performance. Instead it measured a physician's patterns to identify practices experts deemed dangerous or otherwise indefensible. We published our "appropriateness measure" for skin cancer surgery at the surgeon level, and then it was time to press forward. We would soon scale the model. Improving Wisely[5] was about to get a lot bigger.