

**STAVROS** OK, welcome back. We are going to proceed with a POEM lecture. And then the next case will be the full **STAVROPOULOS:** thickness resection. And then there's the lecture and then the colony SD. So we're going to think about per-oral endoscopic myotomy, again, with a focus on the big debate with Heller reflex, what to do about it, and the like.

So let's start. Obviously, this is basic stuff. But rather the [INAUDIBLE] myotomy is really a NOTES version of the Heller myotomy using a tunnel to secure absence of leak. And then the first one was done by Inoue in 2008. We presented that DDW in April of 2009. And six months later, we did the first one outside of Japan.

And now we have still the largest single center published series, I believe. We published *irGE* in July 2017 with 318 POEMs with a particular subset comparing the 48 post-Heller patients with 270 without prior Heller. But this is still, I think, the largest single center Western series. We're about to submit now our update on that with 610 POEMs followed for 10 years. And this is where I'm going I'll be drawing the data from for this presentation largely.

So then our technique has remained relatively stable in the past few years, meaning that we'd go posteriorly at around 5 o'clock if the spine is at 6:00, and the hybrid knife again to enter the tunnel. This is also type 3 patient. This case is from 2013, actually. I still use the video because, as I said, the technique has remained very similar.

So we complete a very long tunnel using the hybrid knife, as you saw today. And then we would do the posterior myotomy full thickness, even since 2013. And then after completion of this long tunnel down to the serosa in the mediastinum pleura, we try not to injure these membranes. Then see at the top, we did live a few longitudinal fibers. It doesn't matter there. And then we suture.

If you want to see more POEMs, seven years of Long Island Lives archived. There are three POEMs every year, one by me, one by Inoue, and one by Zhou with all our particularities demonstrated. And also it demonstrates how each of us change that technique over the past seven years. And it includes generally challenging patients, apart from type 3s post-Heller, pediatric patients, Jackhammer, end-state sigmoid patients, et cetera. So there are good cases that a lot of people starting have used to get themselves into POEM.

Now that said, I'll talk about our series. This is an interesting slide that shows our case accumulation. In this early from 2009 to 2011, in two years, we did about 15. But then it rapidly took off, 2012, '13, '14. Then we hit a stall. From 2014 to about 2017, we hit a bit of a stall because of insurance denials that started in 2014 fueled also by resistance by Heller surgeons.

In fact, a lot of-- a lot of-- a lot of-- this is getting annoying. Can you tell them? Sorry about the intercom.

So basically a lot of that peer reviewers in the US that insurances use when you appeal our surgeons, many-- not even surgeons have any relation to Heller, like bariatric surgeons, or [INAUDIBLE] surgeons, or even thoracic surgeons and I like. So a lot of them have a POEM aversion. So we've got the stall. In fact, we backtracked a little.

But then in 2017, we broke through. I think the people started-- most insurance companies started getting the message. But now we have at least maybe 15% denials. And with careful appeals, we win most of them.

So it has-- in the same period, the technique has evolved. I mean, our technique really hasn't stayed extremely constant in its details for this period. So, for example, the length of myotomy, all expert centers had realized we are doing too long myotomies.

Now interestingly when we started, very, very early, I was following Inoue's lead in doing three to five-centimeter myotomies because of fear mainly of extending into the mediastinum. Once it proved safe, I think we all got a bit trigger happy and moved up to 10, 11, 12, 13. If you look, most of these had mean myotomies in the 11 to 13 range.

But then right around POEM 250 a few years back, we realized that may be too much. We started cutting down, down, down. And now we are more in the five to six-centimeter range for at least the last 100, 200 POEM.

And then what about the real learning curve of POEM? Now with all these cases, it has gotten very, very easy to do a detailed analysis of the operator's learning curve. So you can see if you use POEM duration as a surrogate marker for operator experience and learning, you can see there are multiple so-called plateaus which shows what the fallacy of just using, say, the first plateau, you see after you've done 50 POEMs to declare victory.

So you can see here, we had a plateau at about 100 cases at almost 1 and 1/2 to two hours for a POEM. Then it plateaued at 300 or so cases to something like just over an hour or an hour. And now we are routinely in the 20 minutes to 1/2-hour range. And that's not because we are rushing. It's because you get very efficient at how to deal with things and you get faster and faster naturally.

So you to see there are many plateaus here. So just using one initial plateau, say, up here as the end all, it doesn't make any sense. And this is what happened.

So there were published learning thresholds from, say, Hopkins, Northwestern, and Portland, Oregon when they had only done 36 to 40 cases. I mean, obviously if your whole cohort is 30 or 40 cases, what kind of learning threshold are you going to find? Something less than that by necessity.

So Hopkins, I think, was 13 cases as a threshold. But if it takes 250 cases to get competent at the RCP, I really don't know how 13 cases will make you competent in a POEM, especially if you don't do a lot of ESDs. And then Northwestern and Portland was also in the 20-case range. But again, it's a bit of a fallacy.

Now we published in 2015 60 cases for proficiency somewhere-- there, where you can see probably the first plateau happened, this one, at that 1 and 1/2 hours. In fact, our median procedure time then was, I think, 70 minutes. So it was really the first plateau.

And then interestingly, the operator with the highest POEM volume in the world, Ping-Hong Zhou in Shanghai, published a hundred cases as the threshold, which again, appears reasonable, at least for competence or proficiency. So interestingly, the centers with the biggest volumes would publish the higher thresholds. And the centers that had accumulated lower volumes were publishing lower numbers, which doesn't make any sense because high volume centers should go through the learning curve faster presumably, for variance reasons.

Anyway, so you can see there's constant learning. I mean, now this is a real plateau. I don't think you can-- there's any way you can get a POEM to less than 20 to 30 minutes on average.

Now another way to look at learning curve is to look at the mucosal injury rate. But this had its up and downs, as you see here. And I'll tell you what happened.

So very quickly, within 100 cases to 150, the mucosal injuries dropped to near zero. But then what happened? We started going up again. And that's because we started our anterior-posterior randomized trial with 250 cases. So all right about 150 to 200 to about 450 or so was the anterior-posterior trial.

And clearly, anterior POEM has a higher rate of mucosal injuries than posterior POEM. Other studies than our own, including the random [INAUDIBLE] from India, has shown this. And it makes sense because of the mechanics of getting into the [INAUDIBLE] anteriorly versus posteriorly.

So you can see there was a bit of a bump on the road. But then once we got good at it, it dropped again. Once we got good at the anterior POEM, it dropped again. But now we are back to our posterior ways, and it's now has been zero for the last 100 POEMs.

So there's a learning curve, too. And it's not 23 cases or 18 cases. It's more like 100, 200, whatever. I'm not sure what this was showing, but OK.

So now let's move to the data on the outcomes. So, again, I will draw from our 610 analysis that includes all the POEMs I did, consecutive POEMs from my very first in October 2009 until October 2019, so 610 POEMs. We submit this-- by the way, I did a double in 2020, and it was selected as a poster.

So mean follow-up, 35 months. Mean age, much older than the Asian series. The mean age is in the 30s, like in Shanghai series, I think it's 38, 55 Western series, much older patients. 11% older than 80, 2% older than 90.

You can see the BMI. We get obese patients, even in the achalasia group in America, BMI more than 30, a quarter of the series, slightly more than a quarter. And this goes with the old people, an ASA class 3 or 4 in 1/3 of the patients, so a lot of comorbidity.

Also duration of disease more than 10 years, these can be very difficult POEMs, slightly over a quarter. So these are highly manipulated patients with dilated esophagus and whatnot. That also shows on the stage where approximately one out of five where end stage with sigmoidization.

Type III, as I said, can be difficult, too. I put in red basically things that make POEM more challenging. So type III achalasia at 16%, a little on the high side.

Non-achalasia disorders can be a thorn, as I mentioned in my live case, especially EGJ outflow obstruction. We have managed to keep it to a minimum, being very selective. 4% of our POEMs were EGJ outflow, Jackhammer, and DES.

Prior treatment, half of the patients, including 14% prior-Heller, a quarter prior Botox, a fifth pneumatic dilation, and 3% failed Hellers, mostly from other operators, but a few of our own, too, a very small number.

Now, how does efficacy look long term? Again, there's no point anymore to be looking at three-month, or six-month, or even one-year studies. It's just pointless. There's decent studies with now offering data at 2, 3, 4, and 5 years at least.

So if you look at this data, you can see basically the two- to three-year success rate being in the-- the 78%, these are not better. And the study from three centers, there are a lot of failures for some reason. But about 90% to 100% and 80% to 90% are three years. Then at four years, five years, you get to around 80% to 90%.

Now the two big series, the Yokohama Inoue series with 500 patients and the Shanghai series with 564, had follow-up problems, missing follow-up in 42% of patients are three years, missing follow-up at 37% and 58% of patients at three and four years. So these Kaplan-Meier estimates have a lot of error bars around them. But generally, these are the numbers.

I'll show your numbers in detail. So this is in our upcoming publication, the Kaplan-Meier graph that we'll have in this submission with, you can see here, very well maintained success all the way to nine years of close to 90%. So you can see that one- to six-years' data.

The rest doesn't make as much sense because at seven years, we have 26 people at risk. And the success, mind you, was 90%. And then it doesn't really makes any sense to make estimates for six patients and one patient.

So if you look at the-- the six-year success is 91% on these 64 patients that are over six years old, between 6 and 10. And you can see there's very slow attrition with data that are much better in every POEM. And these are data you can take to the banking and see how small the error bar is because we have only missed 2.3% of follow-ups.

That doesn't mean we don't have follow-up in 2.3% of patients. That means that if one patient missed follow-up at three years but we got follow-up at four years, we count this three-year as a missed follow-up. 2.3% includes that. So courtesy of Rani Modayil, my collaborator, and Maria Kollarus, my nurse, we have even looked at obituaries and every possible source to get follow-up on every single patient.

Now if you compare type I versus II versus III, and we had close to 100 patients that were type III, you can see that the type III is slightly worse using the Eckardt score less than three success criteria. But that can hide some things. For example, the mean Eckardt score for the type III is closer to 1 and 1/2 to two, whereas the mean Eckardt score in the type I and II's is more in the zero to one territory.

So the quality of success-- even though the success, it first looks the same because it's defined within a binary way as an Eckardt score less than three, if you look randomly at the Eckardt score, there is a bit more of a difference. And type III's we know from other treatments are a bit difficult patients. They also have pain, some of them. They have issues.

Now, what about the non-achalasia patients, the non-achalasia batch of disorders. There's an even bigger difference. And again, this is using the binary criteria.

As we look at Eckardt scores, it may be even slightly worse than that. So a clear difference, even though we are highly selective. And only 4% of our series has these patients. If you look at-- and I said what we do with the EGJ outflow obstruction and why it's a thorny diagnosis during my live case.

So if you look at the Cox regression model for risk factors for failure, the only thing that became significant was the non-achalasia disorders versus typical achalasia, with a risk of failure of 3.5 times higher. History of prior treatment, learning curve, type III achalasia, end stage versus then non-end stage, although in positive territory, did not, except for the end stage, didn't reach statistical significance. The prior treatment almost reached statistical significance.

What about adverse events? This is from our series. You can see no mortality, obviously, no need for surgery, or even any IR drains, no leaks, no aborted POEMs, and prolonged hospitalization for more than five days for any reason in the 4%. But this was all mild stuff requiring a few extra days that was not significant.

Delayed bleeding, as I said, we had some suture bleeds where the suture broke and some vessel wasn't adequately cinched. And then they had a delayed bleed from that punctures site that just needed clipping. Then we had two patients that had-- that had to be on anti-coagulation immediately after the POEM.

And then tunnel bleeds without anti-coagulation involved, it was one patient, and one patient readmitted for food infection and an ulcer over the tunnel where the ulcer bled and no treatment was needed. So this is the 1.1% of the bleeds. As I said, no leaks.

There was partial dehiscence of clips when we used to use clips in the beginning of the series of the tunnel that was seen on second look endoscopy but no leak. And we clipped that on 0.3%, two patients.

And then 30-day readmission due to POEM-adverse events are non-POEM-related adverse events, 1.6% and 3.3%, again, mild stuff, C. diff, dehydration, just pain in type III patients that persisted, nothing really requiring anything surgery or IR or anything.

So now the POEM obviously look super safe and looks super effective. But the question is-- and obviously, it gets easier to do as you learn. So now let's compare it to balloon briefly because balloon really has been obliterated by POEM. And then, obviously, the Heller is the main course of this lecture.

So a balloon, we have level 1 evidence, the European randomized trial. Interestingly, it was initially presented at least DDW in 2017 and then published in *JAMA* two years later. But some of the DDW things disappeared in the publication. I included them here because it's curious.

So let's look at success at the one year, 95% in the POEM versus 66% in a balloon, two years, 92% versus 54%. So there's just no comparison. If you look at type III patients-- now, this was in the abstract but did not make it in the publication-- 83% in the POEM versus 33% in the balloon. So the balloon, as we know, it works in type III patients.

Now you may say, OK, fine, yeah, we know it works less, but it's probably safer than a POEM. Not true. Zero severe adverse events in the POEM group, two severe adverse events, including a perforation, in the balloon group, which makes sense.

If you're going to put a huge balloon and rip the sphincter, sometimes you will over rip and sometimes you will under rip. And you really have no good way of controlling that. In the POEM, if you know what you're doing, you can cut the muscle and nothing else. So it makes sense really.

Now if you look at the reflux, interestingly-- although the balloon is not a major issue of contention with regards to reflux. But interestingly, at the DDW presentation, pH studies at one year, 49% positive in the POEM group, 39% positive with a balloon. So if you do a good balloon and rip that sphincter, you should get similar reflux to cutting it with a POEM, which obviously makes sense. Also the GERD-Q score was reported as similar.

Now what did make it in the publication, interestingly, and hopped up on in the abstract was the reflux esophagitis, interestingly, which shows the biggest difference, 41% any reflux esophagitis versus 7%. If we look at severe, meaning more than class A, 10% versus 0%, much less of a difference. You go from 30% less points to 10% less points.

Now, I will discuss this caveat. What does it mean, reflux esophagitis, if you see announcer in a POEM on an endoscopy that is done two months or three months after a POEM? It may not mean reflux esophagitis. But I'll discuss this in more detail.

Now, this is our data. We have been able to get a GERD questionnaire in all but two of this-- all but one of the 610 patients, pretty impressive. And thank you to Dr. Modayil. Grade 2, grade 3, meaning two to four times a week reflux or more than four days a week reflux, was a quarter of the patients.

Now if you look at the pH study, and again, courtesy of Dr. Modayil, we have been able to drag back 2/3 of our patients for a BRAVO. And we get a positive study in 57% of those. But we're not adjusting for stasis and fermentation. And I'll-- and at least a third of these positives could be due to false positives, but I'm going to discuss this in detail.

Now if you look at reflux esophagitis, again, a caveat there that I will discuss, we have been able to endoscope 69% of all patients. And reflux esophagitis, as you can see here, was seen in 49%, mostly class A, 33%, and very little class C, and no class D detected. But again, we have to discuss this. We do the endoscopy at about three months. Again, maybe doing it later would make more sense, as I'll show you.

So, how do other studies line up? Now, these are studies specifically selected to be only on Western patients because Asian patients have less reflux all across the board, whether achalasia, non-achalasia, or any other condition. So Western series that have tried to collect symptoms, esophagitis data, and pH studies, all three I think are the quality studies you need to look at.

So if you look at symptoms, 13% to 28% had positive symptoms, esophagitis, 20% to 49%, and positive pH study, 38% to 57%. We have the highest number. But again, we are the ones that have done the most testing.

I mean, look at these numbers here. Like, for example, the Cleveland group only tested 36 people. The biggest study would be the [INAUDIBLE] with 103. So we are very-- and that's interestingly shows also a high rate of 50%. So maybe the less you test, the less you find.

So what are the caveats here that I was alluding to? First, symptoms, we know what the problem with doing a symptom questionnaire is. And a lot of symptoms depend on just a questionnaire.

And interestingly, and more importantly, a lot of the older Heller studies, that show very little reflux allegedly, depended on the surgeon's nurse retrospectively collecting GERD symptom questionnaires. Now we know that they massively under-represent reflux because 50% of patients with objective reflux have no symptoms. This has been shown repeatedly by many studies.

And also conversely, you get people who have heartburn, chest pain, regurgitation. And these may be actually symptoms of poor emptying and maybe a badly done POEM. Some of the symptoms of achalasia overlap symptoms of reflux, and they are misinterpreted by patient and by physicians often.

Now if you look at erosive esophagitis, there's a [INAUDIBLE] the problem here. Look at this ulcer here, right over the 5 o'clock position of the POEM tunnel. So you're just not going to dissect one-third of the circumference of the esophagus, obliterating all the blood vessels feeding the mucosa, and then you fill the space under mucosa with scar tissue.

This mucosa is going to be susceptible to injury. It's a vulnerable ischemic mucosa. So it wouldn't be surprising if you get a solitary ulcer there, even with a reflux exposure of 3% or 4% or something minimal. So the whole concept of erosive esophagitis has to be put in perspective.

Now, there's no-- if you start not counting these ulcers or manipulating them, you could get spurious data. So it's not clear how to distinguish this from that. And that's why pH studies are more important than, I think, erosive esophagitis. Although, they do have the potential for error there, too.

So anyway, one possibility also, instead of doing the repeat endoscopy at three months or six months, maybe do it in a year and keep that patient on PPIs until then. And this is something that we may consider moving to.

There is obviously selection bias. If you only test selectively with pH testing on endoscopy patients that are troublemakers and are complaining, clearly you're going to overestimate GERD because you're going to be testing a group that is enriched for patients that have GERD. So there's a selection bias.

And then finally when you do a pH study, and Steve DeMeester, when I went to proctor him in USC in 2012 alerted me to that, obviously, he's the son of the Tom DeMeester who made the score, I started complaining to him about this big reflux scores and DeMeester scores on the pH side as I was doing then in 2012. And he said, well, DeMeester scores can be influenced by stasis and fermentation. And I found out Tom DeMeester identified this himself in a publication that I will review. So you have to really look at everything with a critical eye. So let me give you an example.

This was case number 33 in 2012. This was probably one of the cases I was complaining to him about back then. So you can see here total time pH less than 4, 4.7%, barely a positive study.

But look at the reflux episodes. There's really one big one, not those spikes that you see usually. There is an initial spike. But this acid blows then hangs out for hours while the patient is sleeping. That green area is the supine position sleeping on the journal.

So he eats the dinner, like right here on the yellow. Then as he was told, a few hours later, goes to bed and then gets a nasty episode that hangs out for, who knows, two or three hours, giving immediately a positive DeMeester score. And mind you, most of the time, it's between 3 and 4, not really 1 or 2, which is what causes most of the mucosal injury.

So it's basically pepsin activation. There are studies that show that really the critical factor may not be 4. It may be more like 3. Less than 3 is where you get the real injury.

So if you look at this, you're like, well, OK, how do you correct for that? We tried to correct it because-- I'll show you why. But we encountered recriminations from other colleagues. Anyway.

So this is the problem. If you discount anything but the acute episode here, because this may be due to fermentation or stasis or both, you get the pH less than 4, 1.6%, with, obviously, no significant reflux. So this is the issue I'm talking about.

So Tom DeMeester back in '97 identified this. He did fermentation experiments. And for those of you out there that will steal my slide here, please give me some credit on occasion because that happens a lot, I have to say.

So anyway, he did this fermentation experiments looking at pH less than 4 in-- he put food and fermented it at 37 degrees Celsius on a Petri dish. And that fermentation and the lactic acid he created, he got a pH less than 4 in a third of the samples ex-vivo.

Then he used 12 of his Heller patients, 50% of whom had abnormal pH after the Heller, and then he showed that half of those half had a very slow steady drift to pH less than 4 consistent with stasis and fermentation, like what I showed you in our POEM patient. So he said, these are probably false positives. That don't mean real reflux. So maybe the real reflux is 25%.

And he said, use of pH 3 as a threshold clearly distinguished these patients from patients with true reflux since the patients with reflux all had an abnormal percentage of time below pH 3. So he said, we should use pH less than 3 rather than less than 4.

We did a study to look at that. So partially in response to me, this is 2016 DDW. In DDW 2015, when I said that we, as Tom DeMeester recommended, we looked at our BRAVOs to try to separate stasis and fermentation from true reflux, I was accused in the middle of a large DDW room while on the podium by Jon Richter of maybe trying to fudge the data.

So at that point, I think it's still valuable to do that. But at least from the point of generalizability between studies, maybe we shouldn't correct. So some of the pH studies I showed you are probably false positives. But we stopped correcting it.

But just out of curiosity, we said, what if we use DeMeester's recommend analysis using pH less than 3 as the threshold instead of pH less than 4, how would that work? So we got-- we had then 117 BRAVOs. 91 had to be done in our institution. So we had all the raw data.

So we took that raw data in a thumb drive and send them to Parkman at Temple, obviously a motility and BRAVO maven, to interpret the raw data himself, both with the standard analysis, pH less than 4, and with a pH less than 3 analysis. And there are normal data published on what's normal for pH less than 3. It's about 1.9% of the time the pH should be less than 3.

So based on this normative data and this analysis, we found out that 33% of patients had reflux based on the pH less than 3 analogies that Tom DeMeester himself recommended for achalasia. If you use the standard analysis, 55%. And I'll show our most recent data, so pretty consistent.



And what about the-- so, what is right? Well, if we look at mucosal injury with erosive esophagitis, which one predicted best erosive esophagitis? Well, that is clearly the pH less than 3 analysis with three quarters of them having erosive esophagitis, whereas if you look at the pH less than 4 analysis, barely it's a flip of a coin whether this patient will have erosive esophagitis.

However, as it turns out, convincing companies to introduce an algorithm of pH less than 3 in the BRAVO machine is like asking for the moon. And this analysis of the fanciful is difficult to apply clinically, unless you do it manually like we did it with Henry. So these are the caveats there.

So now that we have a good understanding about their reflux caveats and issues, let's look at Heller versus POEM in terms of GERD and efficacy. We'll review this data.

So the surgeons are getting pretty entrenched and combative about this debate. And they come out with T-shirts like this, esophageal POEM, nice little POEM, that doesn't rhyme, by the way.

Once there was a man with dysphagia. His surgeon diagnosed achalasia. He performed a Heller. The patient swallow better-- I mean, how does this matter? I don't understand-- because POEMs cause Barrett's metaplasia.

So-- yeah. And I also have this T-shirt, Make Heller Great Again, obviously a saying of our president. And I hid this face of the surgeon here so he's not embarrassed years from now when POEM has won.

So now let's seriously look at data. So well, the lowest evidence thing is these dirty meta-analysis that are popping up like mushrooms everywhere. I think the best done is the Repici one where they looked at, for example, for pH studies, they tried to do good quality studies. They did an analysis for publication bias, heterogeneity analysis. This is very well done.

So if you look at POEM positive pH studies, 39%. If you remove the Shiwaku Japanese any study because it's heterogeneous and having very low reflux, it goes up to 47%. If you look at Heller, Lap-Heller, again, 16%.

But there's a lot of poor studies. If corrected for publication bias by removing the bias studies, it gets to 22%. So either way, there is a real difference, at least in the first year, of somewhere between 15% to 20%, I would say.

Another meta-analysis that I like-- there's another meta-analysis by Schlottmann, which is not so well done because it includes hundreds, I think, of, if I remember correctly, achalasia Heller studies from single centers with small volumes in Europe and everywhere else that grossly underestimate reflux. I'm talking about reflux rates less than 10%, which obviously are science fiction.

Or actually if you do have very tight Dor or Toupet are not science fiction. But then you get terrible efficacy results. And if you only look at the reflux, it will look great. But you have to always look at efficacy with the reflux because you can make your reflux data perfect by doing a very tight DeMeester, but you will going pay for it on the Eckardt score. So you should really report both.

So this meta-analysis I like because they looked at centers that published comparisons between POEM and Heller from the same operators. These are usually surgical centers where the surgeons used to be heavy Heller surgeons. And when they switched to POEM, they compared their first 30 or 40 POEMs to their most recent 50, 60 Hellers. Not a fair comparison because these people were experts in Heller and they we're just learning POEM.

But nevertheless, POEM had the significant lower Eckardt score than Heller, so it's better. A lot of meta-analysis, including the Schlottmann, won't have shown clearly that at two years even there's a differential in efficacy, not huge, but there's a clear differential, with the POEM being more efficacious.

There was a trend for the POEM being faster. This is because this was learning-- this was the first 30, 40 POEMs of a surgeon. Clearly POEM is faster than a Heller. But this is affected by the studies included.

Trend for the POEM having a shorter hospital stay, it makes sense. Now look at the reflux. Reflux scores similar across the board. These are POEMs and Hellers by the same groups in the same hospital comparisons, so a pretty good way to determine the meta-analysis. Esophagitis in POEM, more didn't quite reach statistical significance. Again, you have to think about how much of this is real erosive esophagitis.

And then pH studies, two studies looked at pH, very little difference. Did not reach statistical significance. For example, in the Swanstrom study, it was 39% positive studies versus 30% in favor of Heller, but by only a 10 percentage-point difference and 9%.

Now, the next level of evidence would be case control. You're in the same center, you try to match as best you can your POEMs with similar Hellers and look at the difference. So this is a case control from this surgical group. And you look at the scores, the GERD scores, not huge difference, not significant. If you look at erosive esophagitis, 32% versus 54%, with about half and three quarters of the patients tested with endoscopy.

And then pH study, they checked about a third in each group. And they found 50% positive pH studies in the POEM versus 30% in the Heller. Again, you see this about 20 percentage-point difference acutely in the first month after these procedures. And I'll get back to that.

Now, there are some case control studies that are a bit irritating to me because there are all kinds of biases. And they're quoted to me repeatedly as evidence of how bad POEM is with nobody really reading beyond the abstract to see everything that is hiding in the methods.

So this is a case control study that matched 31 POEMs with 88 Hellers in a surgical group that, again, went from Hellers to POEMs. And they found amazingly an abnormal DeMeester score in 55% of the POEMs and 17% of the Heller plus Dor, like almost a 40 percentage-point difference. So immediately you know something is afoot here.

So when you look at the details, everything becomes obvious. The pH study was performed at two months when the POEM is not even healed yet and is at its most open and the Heller probably still has edema and it's at its tightest, problem number one.

Then you look at selection bias. Of the 66 POEMs that they performed in this two-year period in that center, only 31 had the pH study, meaning less than half. Clearly for the POEM group, they properly tested the troublemakers and reaching the mix with reflux.

On the Heller group, based on what they say in their methods, from the patients that were included in the case control study, out of the 93 Hellers that they matched to the POEMs, 88 had been pH studies, 95%, which means they were testing everyone, and thus, a lot of patients that did not have reflux. So a big source of potential selection bias.

Then you're like, OK, is that all? No, it's not all. If you look at the prior treatment-- imagine, this is a case control study. So shouldn't you control for prior treatments? 71% of the POEM patients had prior treatment versus 44% of the Heller patients.

And look at the prior Hellers. One-third of the POEM group had a previous Heller, meaning another cut on their sphincter. And 0% in the Heller group had, obviously, a previous Heller. Prior pneumatic dilation, over a third in the POEM group, 18% in the Heller group, again, disrupting the sphincter. So clearly the POEM group had a much more disrupted sphincter than the Heller group.

Duration of disease, very important. Three years in the POEM, two years in the Heller. I mean, it makes you wonder, this case control, what kind of matching methodology was used here?

Finally, you're like, OK, so the reflux was low on the Heller group, OK. That may imply that you guys are making a very tight fundoplication. So maybe you paid for it by having a lower efficacy than the POEM group.

If you tried to find efficacy data in this study, you'll be disappointed. There's no Eckardt score. I don't believe there is even clinical success in the two groups. You can get the hint about that by looking-- they do mention regurgitation, obviously because they consider it he has a reflux symptom, not does it persistent dysphagia symptom.

And interestingly, even though they argue that POEM had a lot more reflux than Heller, there was more regurgitation in the Heller group, 10% regurgitation in the Heller group, 0% in the POEM, making you think that maybe the Heller group has a very tight fundoplication and things get stuck there. Now that's what I have to say about being alert on case control studies about the methodology.

Now let's get to the highest level of evidence, which is the European randomized trial on POEM versus Heller. They compared 112 POEMs to 109 Hellers. There are some caveats here, too. The enrollment was 2012 and 2015 when every European center was at its infancy in terms of POEM technique.

Eight centers had enrolled patients, but most of them were in the top three-- that were in the top three-- they were enrolled by the top three centers. The remaining five centers performed less than 20 POEMs per year. So in 2012 and 2015, the experience in those centers is arguable.

To try and safeguard against that, the study investigators required four to five observed POEMs and 8 to 10 run-in POEMs, very low numbers. I mean, I showed you the learning curve data. This is a bare minimum requirement. And I'm not sure if that's enough to do a randomized trial against Heller with experience Heller operators.

They also pick a very low challenge population. 64% are treatment naive, only 9% type III achalasia. And no patients with prior Heller were allowed to be enrolled. And no patients with tortuous esophagus or dilation beyond seven centimeters were allowed to be enrolled. So a very [INAUDIBLE] group to do POEM on.

Nevertheless, pretty amazingly, the success at the one year was 85%. That's about 10 percentage points lower than what it should be in this kind of group of patients. Lap Heller interestingly, was also 83%, so interesting. Fortunately, they even themselves out.

At two years, it's nearly the same, which means that this is all technical failures. All the failures happened in the first year, and then they stopped-- it stopped getting worse on the second year because now you are left with the successful patients. So these are early failures, meaning inexperience of the operator.

Now, success in type III patients at two years was 83% on the POEM, 78% of the Heller, not as big a difference as you would expect, is it? As I said, this may have been overblown frankly. So thankfully for us, it made it in all the guidelines, and it helps with insurance companies. But whether it's reality, it's a different story. You might not need as long a myotomy as people think.

So serious adverse events, 3% in the POEM, 8% in the Heller, including three perforations and two cases where the patient had to go back to the OR. It didn't reach statistical significance, but it makes sense. I mean, POEM is really safer.

Now if you look at the reflux stuff, this gets really interesting. Mean acid exposure at three months, 7.1 by 6.7. I mean, whoa, this is identical acid exposure. A two years, same thing, 5.7, 5.4.

Now, if you look at the percent that had-- the percent of patients that had abnormal studies, 44% versus 33% in the first year, only 10 percentage points. And then that evens out to both 30% at two years. A patient with abnormal DeMeester score at three months and two years, 49% versus 37% in the first year. But see, as the Dor loosens and the POEM heals and gets tighter, it obliterates the difference, 39% of patients and 34% of patients, matching, for example, the data from Swanstrom's study of comparing his POEM to his Hellers that I reviewed on the meta-analysis that showed the centers that compared their POEMs to their Hellers.

Now, if you look at severe esophagitis at three months and two years, you can see it goes down at two years because of healing of those ulcers, whereas the Heller one goes up from 20% to 29%. The net difference for overall is 44% versus 29%. But if you look at severe C/D esophagitis, it goes numerically higher in the Heller group than the POEM group, 6% versus 5%.

And then overall heartburn symptoms, 41% in both groups, identical. So only the reflux esophagitis reaches statistical significance, and only at three months. And we know what that is because at three months, you have all kinds of ulcers caused by the tunnel.

Now, I keep alluding to the fact that checking out one year or less favors Heller versus POEM on the data. And this is why. These are long-term evolution of their reflux data in Heller plus Dor or Toupet.

And I picked randomized studies to make sure this is high quality data. These are all randomized studies. For example, this one compared Dor to nothing, this one compared-- this one looked at Dor versus-- Heller plus Dor versus balloon. And this one looked at Dor versus Toupet. And this one looked at Dor versus Toupet.

So if you look at one year, you see that the reflux rate is 20% to 40%. That's a prospective blinded randomized trial. So now you can see the real reflux. These are the best centers in the US and the best centers abroad doing a randomized trial to prove whether Dor or Toupet is better.

Interesting, they got conflicting results. But the overall rates are similar in line between 20% and 40%. So this data that my favorite surgeon's quote, that, well, I'll do it there and Heller plus Dor [INAUDIBLE] and you'll have zero reflux, well, that's common, or less than 10% are imaginary data or bad retrospective data.

Now, if you look at this Richards study that checks Dor versus nothing, at 11 years, the people that got no fundoplication, that people that got Dor fundoplication, 69% and 57% had symptoms, 77% and 71% were on PPIs, and the GERD-HQRL was very similar, whether you've got the Dor or nothing.

Now, if you look at that *New England Journal* study comparing balloon and Heller plus Dor, at one year, the reflux on the Dor was 23%. At four years, which was published later in [INAUDIBLE], was 34%, a 50% increase within three years. So it's all consistent

Now you want more data, this is a 30-year study of Heller plus Dor or Toupet, with 67 patients followed over 30 years. And they did pH studies on them, which is remarkable. Positive pH study was in 28% at greater than 10 years and 53% at greater than 20 years. 22% were long-term failures due to reflux. And 13% developed Barrett's.

If you look at this study that looked at 51 patients with laparoscopic Heller plus Toupet, if you look at six years mean, range 1 to 12, 27% had heartburn and 65% were on acid suppression therapy.

So I hope I convinced you that the difference is small at one year and becomes essentially negligible, in my book, at 4, 5, 10 years and beyond. No reason to have an invasive surgery, in my mind anyway.

So PPIs work. They are the cornerstone of our management. Again, notwithstanding poorly done data mining studies that drag 100,000 patients into some computer algorithm, the cardiology randomized study clearly showed that we shouldn't worry about the PPIs too much. Thanks to the cardiologist for doing a good study for us. So PPIs work.

This is from [INAUDIBLE] group where the people go into clinical remission with PPI. And this is the group that has 50% reflux on pH study, and the group with the most pH studies after our own on that little table I show. So very reliable data.

For erosive esophagitis, I do BID PPI for six weeks, and then they go back to QD and they all heal. And they state reasonably that none of the patients with GERD considered their symptoms or the need for continuous medication as a significant problem when compared to their preoperative dysphagia and other symptoms of achalasia. So basically, they are very happy to trade the achalasia for a pill a day.

Now, things are getting better with time on the POEM side. As I said, the myotomy heals and things get better. And PPI works better and better.

Now, that's why I like this study from Mexico. They were able to almost do-- they followed 65 patients. And they had pH studies at one month, six months, one year, two years, three years, and four years on at least 50 of them-- 59, 65, 55. So most of them got pH studies in all this time posts.

And you can see the quality of life in blue-green here. And you can see the clinical score, light orange, and endoscopy erosive esophagitis, class ABCD. I know you can see that here on the darker orange, and this very dark orange, the DeMeester score, pH studies.

So you can see at one month, things looks bad. At six months, they look a little better. And mind you, most of the studies were with people using PPI as needed.

So they don't necessarily show that reflux disappear with time, but it shows that it's extremely manageable with, say, at four years, 5% having an abnormal DeMeester score and 5% having class A esophagitis that may be an ulcer, and no BCD, and 2% have symptoms while on PPI treatment.

And the quality of life remains great. Even though it declines a little bit, it remains very good. So PPI management looks like it's very effective.

Look on it in a different way is, how many patients are disgusted with PPI management and one fundoplication to get off the PPIs? Very few. As I said, I can hardly convince them to have a tiff.

And I'm talking about the most severe ones, quite apart from convincing them to have a Dor or Toupet, which I do offer on the severe ones along with a tiff. Only four people out of 610 took us up on this offer, 0.7%. In the Japanese retrospective multi-center study, 0.1%, 2 out of 2,200, elected to have a fundoplication for reflux after the POEM.

What happened to these four people that got their fundoplication? Not good things. Two have Eckardt scores of 2, not ideal. But all of them are back on PPIs at anywhere between two and three years or so.

This one on BID PPI, actually, after Toupet. And this one QD after partial Nissen and Toupet. And this one who got a Dor, this one had end-stage disease. His symptoms were interpreted as reflux by a surgeon that did a Dor in appropriately.

And after the Dor, he went to esophagectomy, a very sad story. He went to another surgeon. He wouldn't come back to our center because he felt his POEM didn't work. And that's how we were responsible.

Mind you, he had a 15-centimeter sigmoid esophagus. But some surgeon convinced him that it's all reflux, so he got a Dor. But a year later, he needed a esophagectomy. So all of them, not success stories, I would say, overall. So yeah. So this is the landscape over here.

Now, can I blame the surgeons? No. The worst thing you can do on a POEM patient who is doing fantastically on the eating side-- and those three were doing fantastically. One of them was an anesthesiologist from California that looks like an old Olympic athlete that climbs mountains and stuff that didn't want to take PPIs. So, all of them had fantastic results on POEM, those three.

So if you take that patient and you do a tight Toupet and now they can't eat again and they're back where they were before the POEM, they're going to be mad as hell, and you're going to be going back to the OR to take it down and do a looser one. So the smart surgeon opts for a looser rather than tighter fundoplication. And obviously the effect is minimal, as you see.

Now what about-- let's say, OK, fine, so you manage them with PPIs. But is there Barrett's and Barrett's carcinomas all over the place because you guys are not doing a good job with the PPIs? So let's look at long-term sequelae on studies that have longer-term fallout.

So this Northwestern study with mean follow-up of 2.4 years, they got two peptic strictures in people that were non-compliant that responded to dilation. In this three-center study with minimum follow-up of two years, ranging between two and four years, one developed peptic stricture, some non-compliant older patient that left the country that the POEM was done on and was lost to follow-up, and two got a short segment Barrett's over a mean of about three years.

Sorry. In this Teitelbaum study, minimum five-year follow-up. These 29 patients had at least five-year follow-up. One develop short segment Barrett's five years after POEM. So not terrible.

And you have to think about what the background rate of developing Barrett's is in these patients and what's the rate of pre-existing Barrett's that may have been missed by the people that did the POEM. Especially people who had prior Heller that didn't work, prior balloon dilation, they may have developed some Barrett's from the previous reflux that is blamed on the POEM. So you have to be careful about taking the Z line before you do a POEM.

I'll show you why. So we had five patients that develop short segment Barrett's with no dysplasia. Mind you, that's 1%. At two years, four years, five years, five years, and six years, all of them did not want to take their PPIs, refused, were taking them as needed, et cetera.

Now interestingly, more than the people that developed it had it already. 10 patients, or another 1%, had pre-existing Barrett's when we showed them, including one that actually had a non-detected by the referring doctor T1a carcinoma. So we had to do an ESD first on the POEM day and then do the POEM three weeks later. So not much Barrett's.

What about peptic strictures? We had five patients that developed peptic strictures at 3, 4, 4, 5, and 6 years. All responded to dilation, one dilation, two, three, and one of them four dilations. All PPI non-compliance all resolved with dilation PPI. And all these patients, mind you, they all went back on their PPIs because they learned their lesson.

So GERD after POEM appears manageable with PPI. But is there anything more that we can do here? People think-- we discussed this a little on the live case. What about anti-reflux procedures?

This is a MUSE case report from Hopkins, not fantastic. The DeMeester score went from 124 to 62. This is a series from [INAUDIBLE] of five TIFs, not fantastic because he did not report any pH data or GERD scores, just some qualitative data.

But as I said, one, it's a reasonable option for those probably that have severe reflux after POEM, whatever it may be. 5%, 10%, 15% at most that would be. And we use it that way.

There's this fanciful approach. POEM is-- Inoue is innovating again with his POEM plus F, really crazy stuff. These are notes on tier fundoplication. Application. He showed this on our Long Island life course in 2018 immediately after he did his first one in August of 2017.

This may be the first time he showed it in the United States. You can see it if you go at the archived footage from Long Island 2000-- Long Island Live 2017 under Inoue Lecture, you can see this video of his first POEM plus F. And you can see his discussion during the lecture.

So he does an anterior POEM. That's the myotomy. He gets through the myotomy, cuts the peritoneal membrane. That's the back of the liver, so it's relatively safe.

So he cuts the peritoneal membrane here until he makes a hole in it. He can go through, grab the fundus of the stomach, test its mobility with a grasper to see if it's going to come and if he can do the fundoplication. And then once he does establishes this, he puts an endoloop with clip. We can also do freehand suturing or over-stitch suturing on the fundus, one attachment and anchor, multiple clips holding the loop. Then he put the other end of the loop on the edge of the myotomy and then closes the endoloop.

And while he does that, he looks from inside with an XP scope to see how the fundoplication looks. And it looks pretty interesting. So you can see here it's beginning to look more and more like a Dor visually there. Not bad considering how it's made. So very good stuff, but I don't think it's ready for prime time.

The idea of putting endoclips in the peritoneal cavity forever is one thing. The idea of doing a fundoplication with just fixing the fundus of the stomach and the ends of myotomy, I mean, the surgeons take it and fix it to the crura with very nice transmural sutures and anchors and pledgets, and they still have it loosened in a few years. And I showed you the data and how bad they get off their 5, 10 years.

So doing this with endoclips and an endoloop I don't think is doable. He did published data in endoscopy, but again, not with rich pH data or any durability data. But we are awaiting that. I'm intrigued and following closely. But I don't think I could sell this on a Western IRB exactly at its current form unless modified in some way.

Now, what about technique modifications while you do the POEM, doing it in ways that will be less refluxogenic? Some people have looked at full thickness versus circular only, long versus short. I don't have time to review all this data. But there are a few studies, mostly small, mostly retrospective. And the data are weak, equivocal, or negative.

If you look at the orientation, anterior versus posterior, there are more data. You can see here four randomized trials. Ours, as I said, is the largest with 250 patients. There's the International Multi-Center Trial with 150. They reported some late data on that now. This is the original data, but nothing different on the most recent data.

And this has the problem that it's seven different centers with small numbers of patients, people that favor anterior, people that favor posterior, people with very low experience. It's hard to mix all that up and expect to have the power to detect anterior/posterior differences. There's the Chinese with 46 patients and the Indian with 60 patients.

And they found some differences. There's a meta-analysis of them now by Rodriguez de Santiago. And I believe the senior author was Inoue. Looking at them at the meta-analysis of this, unfortunately for all of us, these are disappointing. As I said, not much difference on the reflux.

The only one that showed the big difference in reflux was the Indian one, which is the smallest study or the next smallest, 30 patients in each group. So it's amazing that they found statistically significant differences in reflux. But none of the other studies, including our own, saw a major difference. So similar GERD, similar GERD on endoscopy, similar GERD symptoms.

The difference was less adverse events, mainly less mucusotomies. And posterior was faster-- faster closer, but really overall faster, but statistically significant faster closure. And there was a slightly less length of stay.



There are questions coming in, but I'm going to address them at the end because I'm almost at the end. And we are at 11:15. 10:15, 11:22, one hour and seven minutes. So we have to start thinking about the next case, too.

So anyway, these are the data on anterior/posterior. So it didn't really give us the answer we were looking for reflux. So enter the Anti-Reflux POEM, which, as I said, we started doing in April of 2018.

The goal of the Anti-Reflux POEM was to preserve the important anti-reflux function of this sling fibers of the sphincter. And I'm glad that this is spreading around. I had some observers from Spain a year ago, I think, who are now doing it and actually publishing it, et cetera.

So anyway, this is our Anti-Reflux case control study. 122 patients received this technique from April 2018 to September 2019 after we excluded patients with prior Heller, which would not make sense to include because they would immediately have an anterior myotomy that may or may not include sling fibers, and other foregut surgeries.

So after excluding those, we matched them to conventional POEM. And to avoid any learning effect, we went immediately before April 2018 and included the 200 patients immediately before that time, basically POEM 269 to 468 to exclude learning effects. And we call this the control patients.

And then from those 178 that did not have a previous Heller and those 122, we were able to create 116 pairs that were perfectly matched by propensity score for the six important covariates for reflux-- age, gender, BMI, HRM type, stage, and prior treatment.

I would submit that, mind you, the study I did, a double 2020. And it was also selected as a poster, just saying. So maybe we should have done something better, I guess.

And the reflux versus conventional, dramatic results. I didn't expect such a difference. The positive pH studies these went from 75% to 43%.

Now, this is higher-- in our whole series, it's 57%. But this 75% is also by matching exactly non-prior Hellers-- it's a subgroup of non-prior-Heller patients that were exactly matchable to the anti-reflux patients. That's why the percentage is higher because when we match exactly those patients to those, this group had 75%.

Total acid exposure, 4.1%. The average patient in the anti-reflux group has a normal acid exposure, pretty amazing really. And you can see the IQR between 2 and 6.5. Now on the control group, the IQR is between 5 and 17 and the mean is 10.4, clear major reflux on the average patient.

Now, if you look number of refluxes, 29 versus 53, highly significant. The DeMeester score was better, and the erosive esophagitis was better, but didn't quite reach significance maybe because of all these caveats I mentioned about the DeMeester score being affected by fermentation and stasis, and then all the erosive esophagitis, ulcerations, and whatnot.

GERD symptoms, at least twice a week. 7% in the anti-reflux, 22% in the control. So really excellent data. And did we suffer a worse Eckardt score? Because when you make a POEM that is tighter, you have to worry about, are you going to pay for it in dysphagia?

The Eckardt score is 0 on the anti-reflux, 0 on the control. Although, the IQR, the top 25% worst patients had an Eckardt score of 1, whereas here it was still zero. So there is a small difference.

Hopefully this does not translate into a bigger difference as we follow this patient out to one year, five years, 10 years. But we are following them closely. And percent of patients with follow-up, 100% in both.

If you look at predictors of positive pH study in a multivariate regression, having an anti-reflux versus conventional POEM was the most important, six times less reflux in the conventional group-- in the anti-reflux group versus the conventional group. The male gender, it got to 3.7. And BMI, for every 10% increase-- for every one point increase in BMI, you got a 10% increase in GERD rate, not absolute 10%, 10% relative increase.

So in conclusion, POEM has matured with outstanding two- to five-year outcomes. POEM represents a first line treatment for achalasia that has equivalent or superior efficacy to Heller. And currently, the absolute difference of GERD rates between Heller and POEM is less than 20% at one year and erodes further with time.

GERD after POEM is manageable with PPIs with good patient satisfaction and no evidence of a high rate of complications. And the POEM technique continues to be involved, including refinements that may result in a significant lower GERD rates, including our patented anti-reflux POEM.

And this is what they usually do on August 14. I'm sailing with my brother in Greece, the Greek islands for a week. That is me down there.

The photograph was taken by my crazy brother, who had climbed up the mast to take an overhead picture. But we are in a little isolated cove hanging out and swimming in that beautiful sea there. So that's what I would have been doing. But now I'm going to go to Montauk instead. Not quite the same sea, I have to say.

So that's it. Now, let's answer some questions. Any suggestions-- this question says, any suggestions other than PPI or conversion to gastric bypass for POEM in sleeve gastrectomy patients where reflux is a problem? Well, that's a good question.

Definitely conversion to a bypass is a reasonable choice, especially if the patient is not very thin. These are very difficult situations. You can also try TIF, you can try overstitch with ablation that [INAUDIBLE] has reported to tighten up over there. You can't try the ARMS procedure that Inoue has reported.

But definitely converting is a very reasonable option I think because none of this-- I clearly don't have a huge faith in these endoscopic techniques in terms of keeping a sleeve patient works because reflux is particularly bad, happy for the rest of their lives. So I would maybe be leaning towards a conversion to a rue.

I don't think-- do you see any other-- no other questions so far. Maybe we could address some questions later, too, maybe even during the FTR. Although, the FTR will look a little more concentration on my part on the POEM and the Z-POEM.

So I will start moving towards the FTR. Maybe we can answer any last one, two questions before doing-- before getting thick into the FTR, in which case, maybe we'd better ask questions on that. All right, thank you. I'm going to go move on to the FTR shortly.