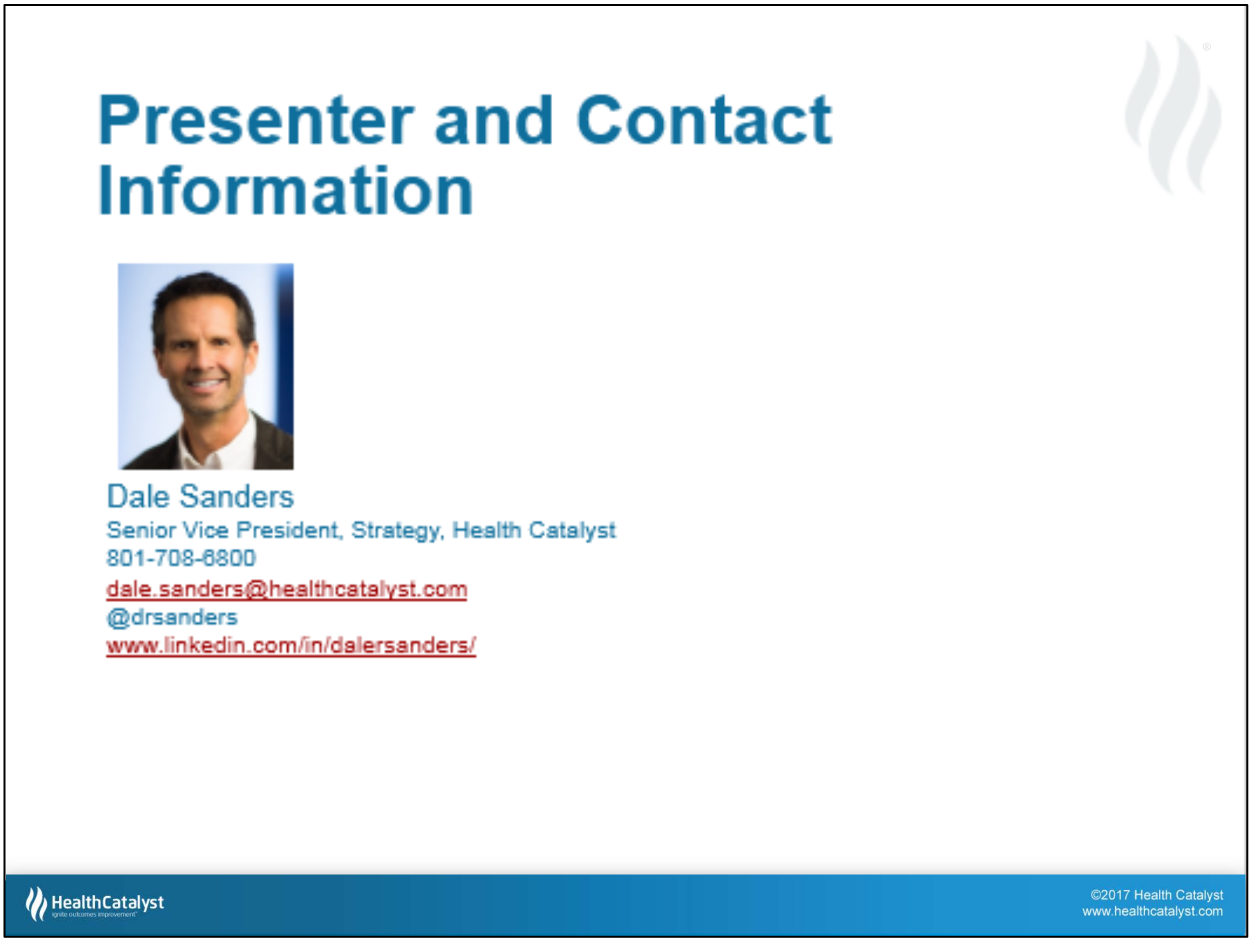
There is a 90% Chance Your Son is Pregnant
Predicting the future of Predictive Analytics

There's a 90% Chance Your Son is Pregnant
Predicting The Future of Predictive Analytics in Healthcare [00:01]

[Dale Sanders]
Thank you, Tyler, and thanks everyone for joining us. Happy holidays to everyone out there and I hope you have a great holiday season with you and your families.

The title today is a little bit of a foreshadowing of my – just a little bit of cynicism associated with predictive analytics in healthcare. It's a little bit overhyped but that said, I'm mostly just having fun with the title and I'm optimistic that we're going to do good things with predictive analytics in healthcare, but I think there are some boundaries and some fences around that and that's what we'll talk about today.

So let's get started…

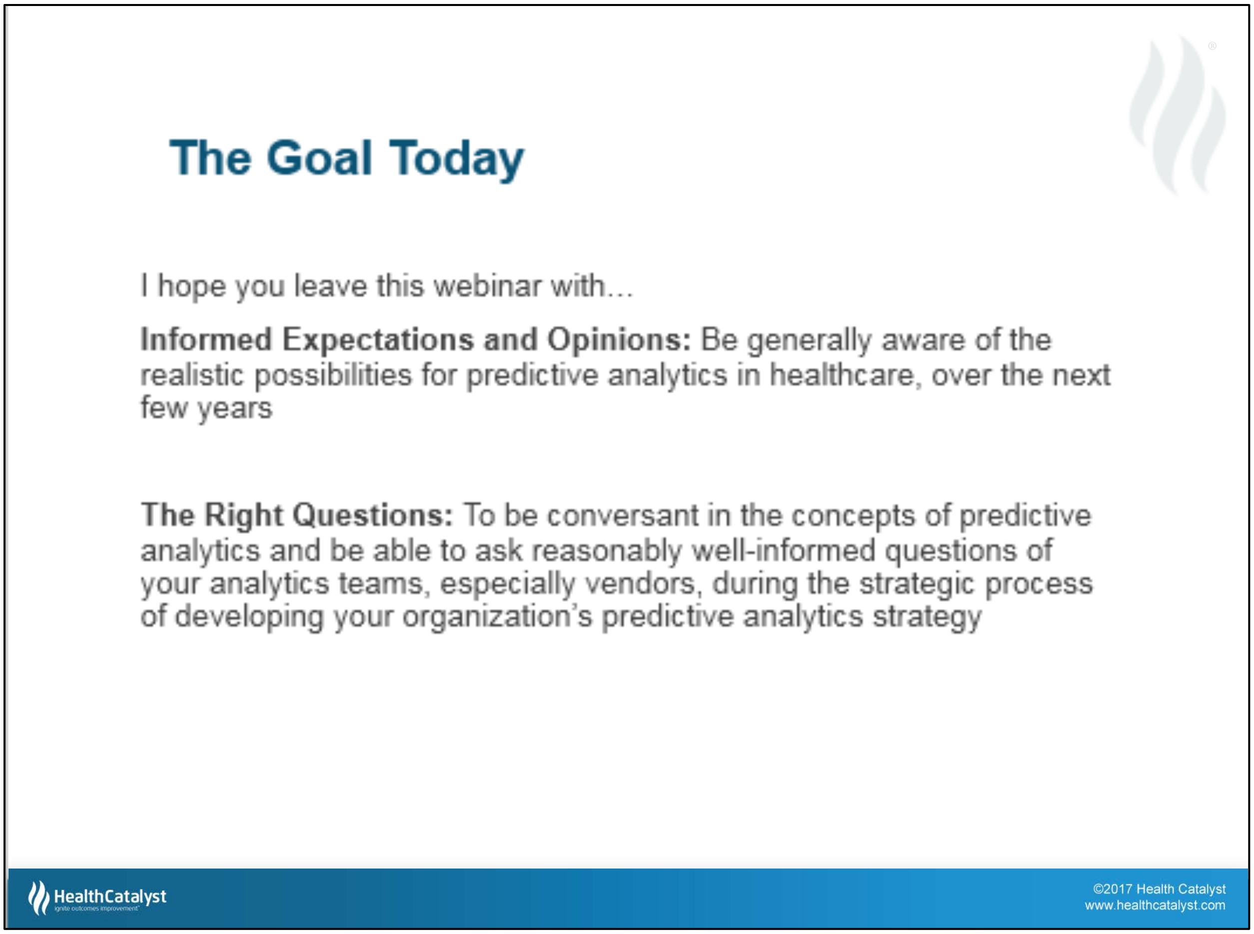


Presenter and Contact Information
Dale Sanders, Senior Vice President, Strategy, Health Catalyst® [00:42]

So there is my contact info and all that sort of thing. If you folks are interested, I'd love to contact you and interact on these topics. Thank you so much for sharing your time with us today.

Acknowledgments [00:57]

I would like to acknowledge key folks, and I'll never be able to put a list together this long enough to do this justice. There are so many people that have been so good to me in my career on this topic, but David Crockett in particular, he is our machine learning and predictive analytics guru, incredibly capable and always helping me learn more about this topic. Dr. Eric Siegel at Columbia probably doesn't understand how influential he has been but I've been a big fan of his writings and his books and articles for many years. And my mentor and dear friend, Ron Gault, who I worked with at TRW, really introduced me to the quantitative side of risk analysis and predictive analytics. And then of course I have to thank Wikipedia too. They have been – you know, it's just still a great source of information about this topic in a variety of ways, and I encourage everyone to take advantage of that.



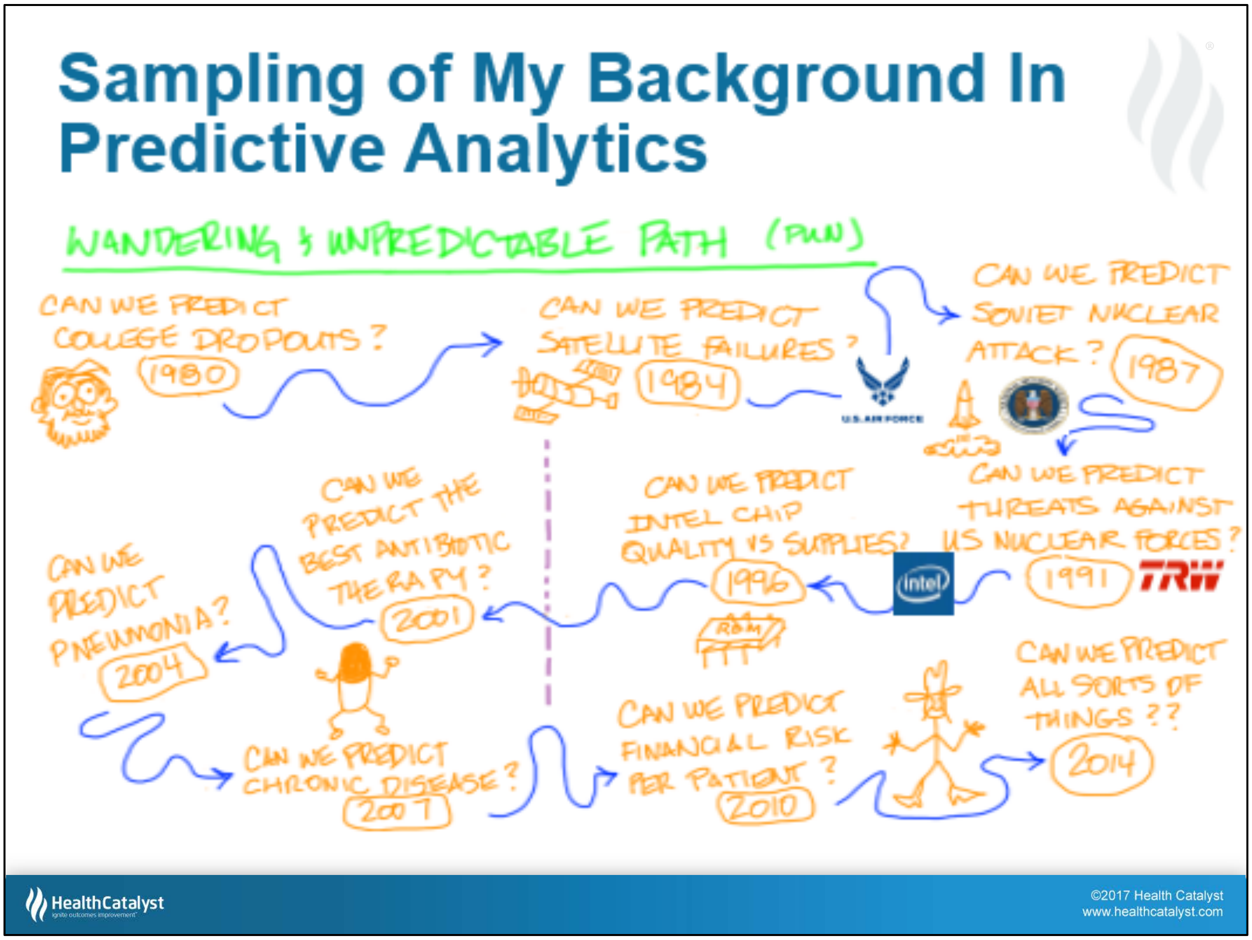
The Goal Today [01:55]

So one of our goals today, there are a couple of goals today, is I'd like you to leave this webinar with an informed set of expectations and opinions, so that you're generally aware of the realistic possibilities of predictive analytics in healthcare over the next few years. And I put that over the next few years because that's kind of a timeframe that I think predictive analytics is going to evolve. And after the next three or four years, as our data ecosystem expands, I think our expectations of predictive analytics are going to increase. But for the next couple of years, I think they're just a little bit overhyped, and we'll talk a little bit more about why that's the case later.

I also want you to be able to ask the right questions. It is kind of odd. My background in this field is not all that technical, and I'll go into that a little bit in greater detail. But I seem to have an ability to ask unusual questions that are reasonably well-informed. Sometimes they're naïve. But those unusual questions tend to reveal sometimes unique and creative answers from people that are generally a lot smarter than I am on this topic. I'll talk a little bit more about that and what that means later and the role of folks that are more typical that don't have a predictive analytics background, what kind of role you can play in this space going forward. So, being able to ask direct questions of your vendors, as well as your teams is one of my goals here today as well.

Agenda [03:36]

The agenda, we'll talk about basic concepts and fundamental assertions; predictive analytics outside healthcare, which I think is interesting and in some ways is predictive of what may happen in healthcare as well; give some examples of what is going on inside healthcare; and then those key questions to ask vendors and your analytics teams.



Sampling of My Background in Predictive Analytics [03:57]

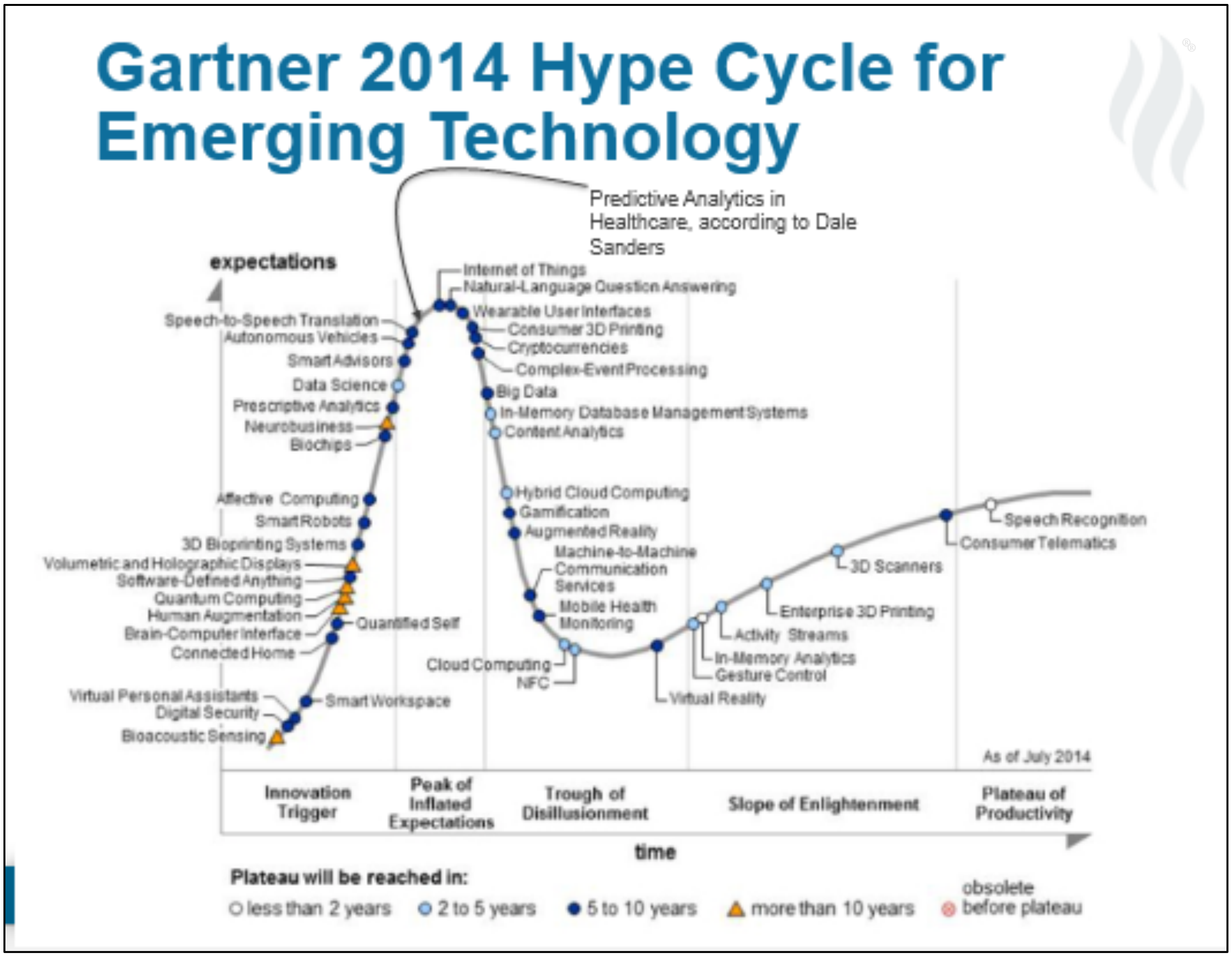
So this is my cartoon drawing of my career as it relates to predictive analytics and I would say it's been incredibly unpredictable. My first exposure to this concept was in 1980 when the four of us got together to program an application for our Computer Science 201 class to see we could predict who would be most likely to drop out and as it turned out it was a pretty accurate model. But boy, looking back, it was very unsophisticated and lots of manual manipulations from that.

In 1983, I joined the Air Force and I spent half of my career, 8-year career, in the Air Force in space operations and I spent the other half in nuclear warfare operations. And at that time I was a customer of predictive analytics, trying to predict satellite failures and satellite problems. Then in the later stages of the 80's when I transitioned into nuclear warfare operations, we're all about trying to predict whether or so the nuclear attack would occur or not. I got out of the Air Force in the early 90's and did some work for the National Security Agency and TRW, trying to predict threats against nuclear forces from terrorists and other organized groups. I went on to work for Intel where we were trying to predict the relationship between the quality of the chips in the fabrication plants and the supply chain and that turned out to be maybe the most significant financial impact that I've ever seen certainly in my career from predictive analytics. Tens and millions of dollars resulted from that.

I crossed the boundary into healthcare in the late 1900's and early 2000 and went to work for Intermountain, and one of the more compelling things that we did there was predictive analytics. We did a lot of things actually but the more compelling example was the Antibiotic Assistant that Dave Classen and Scott Evans and others were heavily involved in, and I got to watch that evolve and I got to watch that operate within our hospitals and learned a lot from what you can do with predictive analytics if you have the right people and the right data and also I might note the right user interface back to the point of care in the electronic medical record. We also worked on predicting pneumonia, predicting chronic disease.

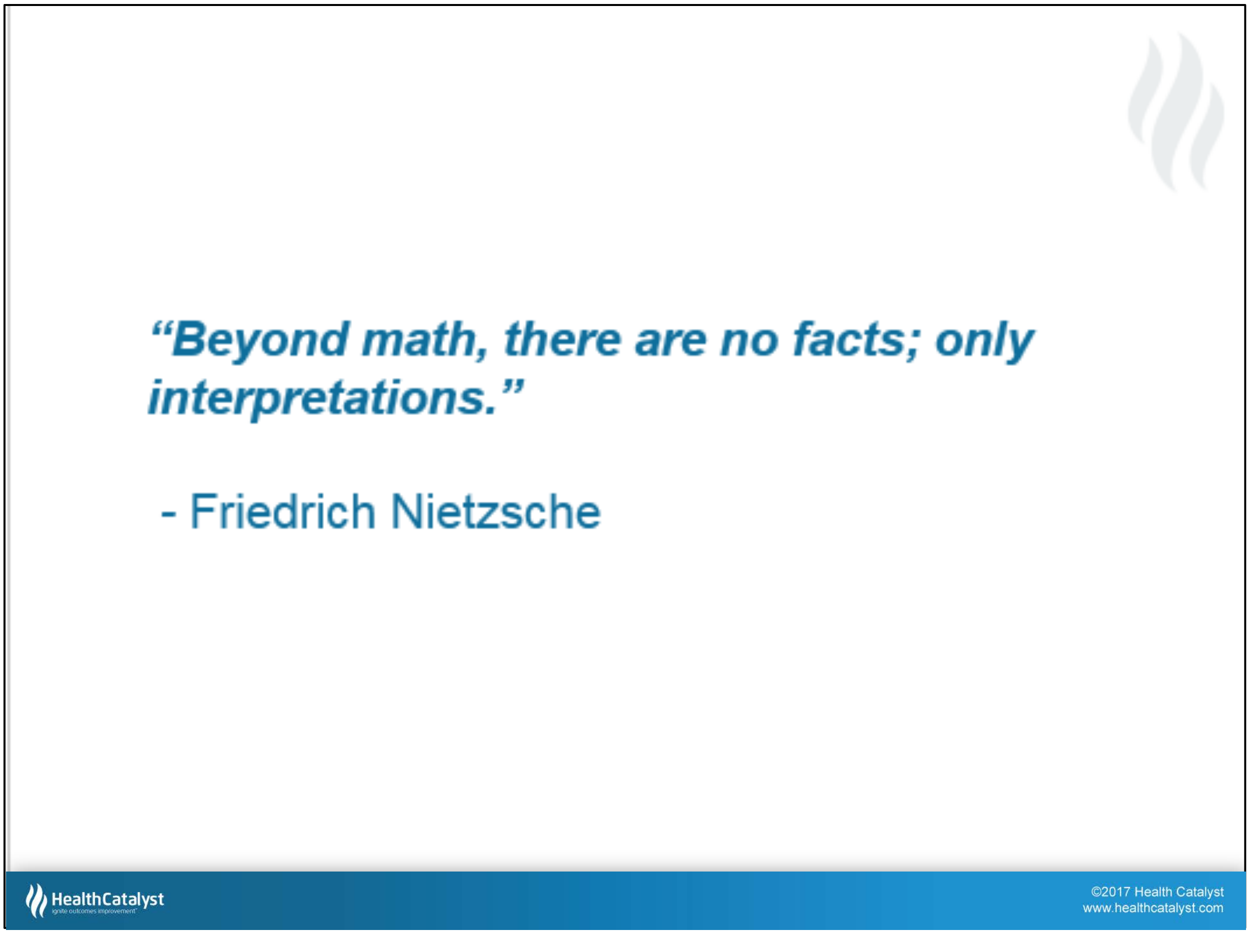
And then I transitioned back in 2010. I went down the Cayman Islands which was a true public health system and a true accountable care organization really. I was the CIO there and at that point we were trying to predict all sorts of things but particularly financial risk on a per-patient basis. And I might mention that, you know, in this world of predictive analytics in healthcare, there are two forms of predictions that we have to start embracing, I think, a little more clearly. Even though most of the time we focus on clinical risk, financial risk is becoming a bigger and bigger issue. And so, predicting financial risk and predicting clinical risk are actually two different things that can be merged together and they certainly have a relationship, but we need to start thinking about those somewhat separately certainly because the data and the pre-processing of that data is considerably different for both.

And now here we are in 2014 when we're trying to predict all sorts of things. That little cowboy there is my indication of the Wild West. Predictive analytics is highly hyped right now and we're looking forward to what may happen in the future. It's going to be interesting.



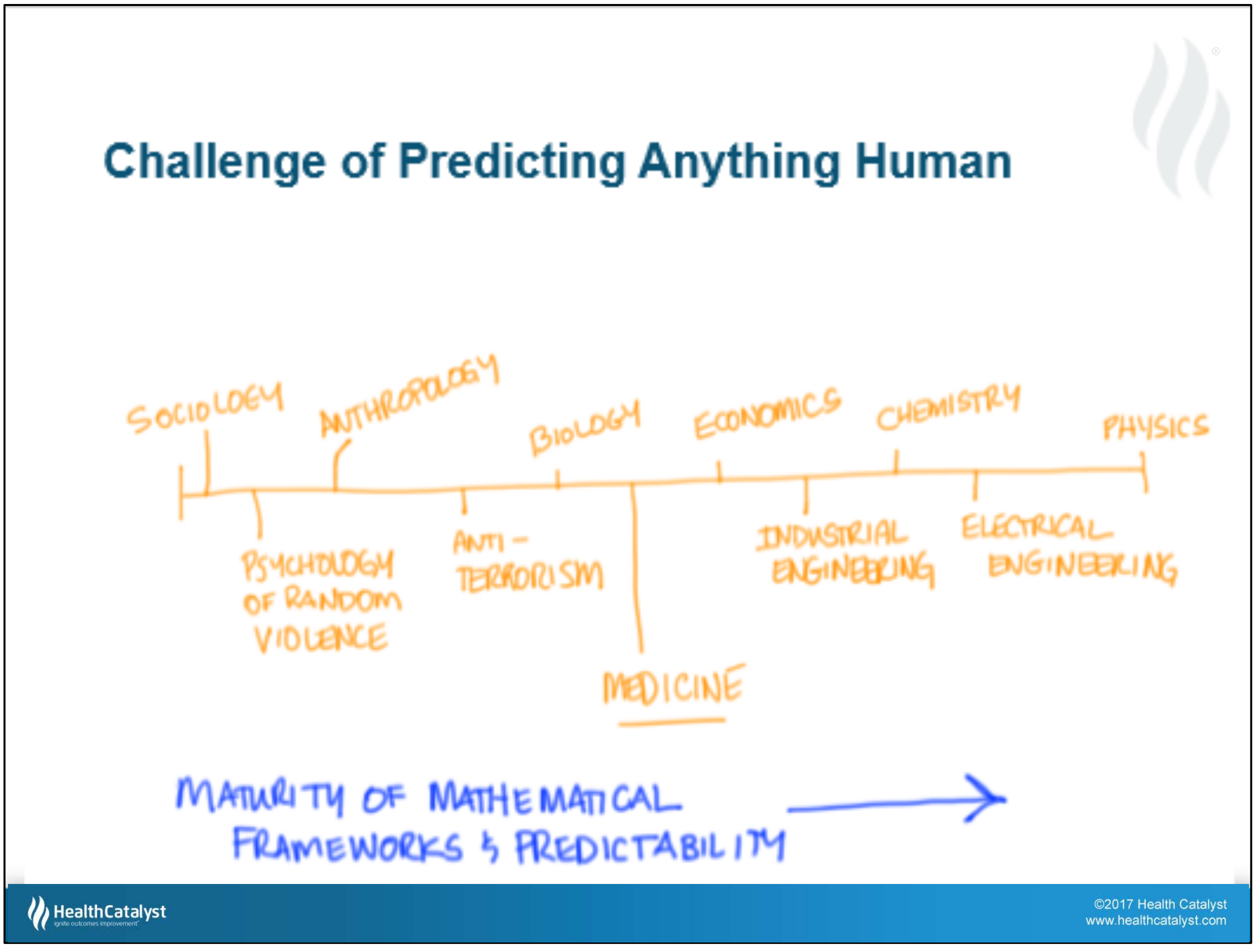
Gartner 2014 Hype Cycle for Emerging Technology [07:55]

So speaking of hype, I borrowed this from my dear friends that we had a conversation this morning with Laura Craft and Vi Shaffer at Gartner, our usual checkup with them and usually very informative. Predictive analytics doesn't show up specifically on the Gartner Hype Cycle but I placed it near the top of the hype cycle right now with prescriptive analytics kind of on the upswing. As we get into this, we're going to find that it's not as easy to implement these predictive algorithms and processes as we hope and it's certainly not as easy as a lot of the writing and the press would indicate, but there's still plenty of opportunity. And again, my hope today is that we'll talk about what those realistic opportunities are.



“Beyond math, there are no facts; only interpretations.”
- Friedrich Nietzsche [08:55]

So I love this quote from Nietzsche about the kind of the fuzziness of humans, as how I look at this, and that is that beyond math, beyond 2 plus 2 and 3 minus 1, there are no facts; there are only interpretations of the world around us.



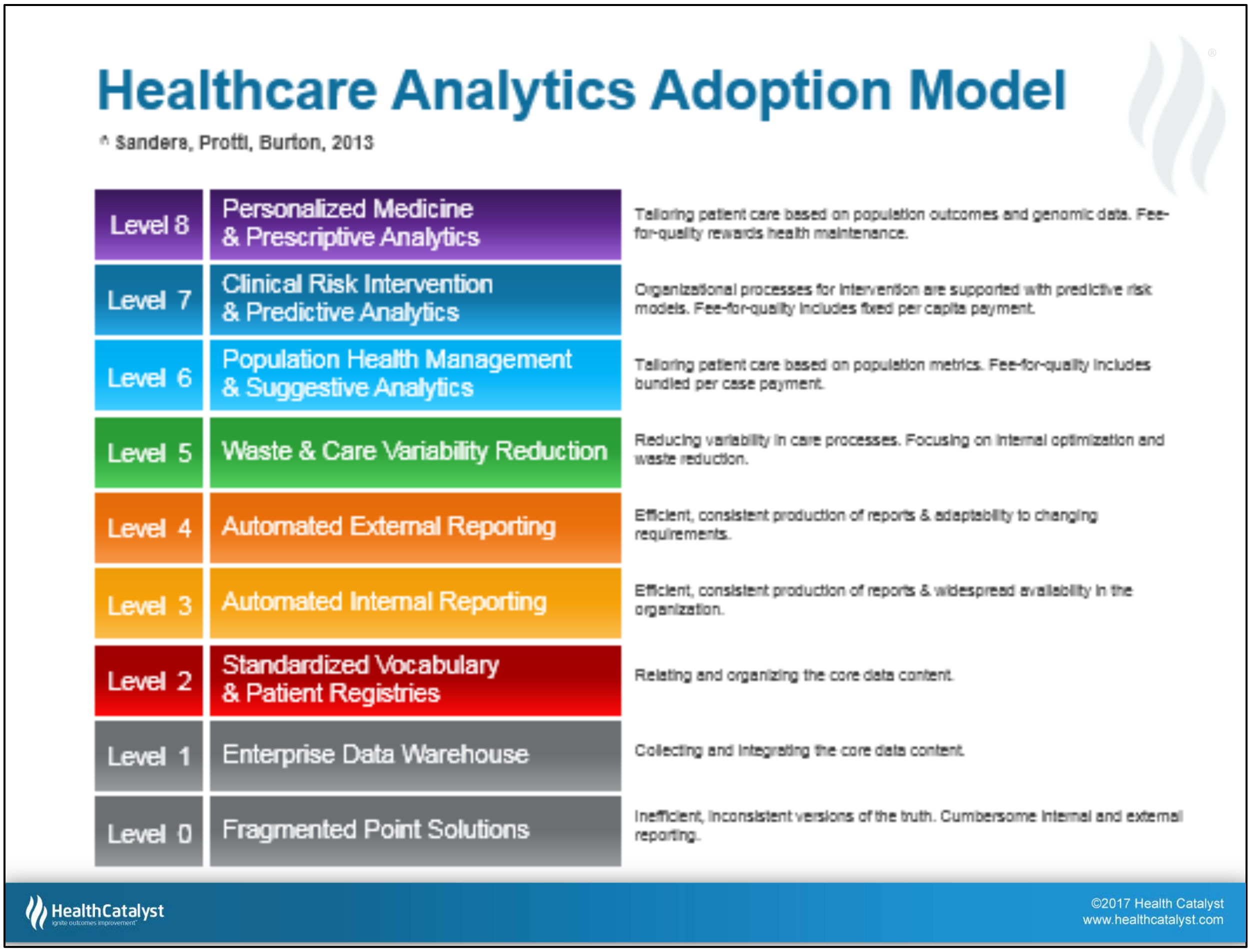
Challenge of Predicting Anything Human [09:12]

And if you look at the progression of predictability from left to right on this scale, the socio-environmental areas of life, the psychology, anthropology, where our mathematical models haven't matured significantly are generally the most difficult to predict, and I can speak first hand from my work at NSA and TRW when we were focusing on the anti-terrorism activities. Really really hard to predict who is going to do what and when. That said, there are some interesting parallels between anti-terrorism and healthcare. Both use registries and as it turns out, there are quite a number of overlapping socio-economic behaviors that drive people towards terrorism that also have an effect on the poor healthcare outcomes.

The medicine is kind of in the middle of all of this in terms of our mathematical maturity and our modeling. Physics clearly on the right is the most predictable that's why we can launch a space probe and land it on an asteroid, it's millions of miles away. That's what predictability is all about. So we still have ways to go with the predictive models and the maturity of the mathematical models in medicine and this is much of a philosophical problem as it is a technical problem.

What Should We Expect in Healthcare? [10:48]

Gary King at Harvard University is the Director of the Institute for Quantitative Social Science. And you know, the takeaway that I bring with me from my experiences in this world of predictive analytics is that machines are predictable but humans aren't. And Dr. King essentially said the same thing. He says that people are influenced by their environment in many different ways and these environments change more quickly than people do, and all of those rolled together into variables that are very unpredictable and how they will impact a person is even less predictable. So given the same situation tomorrow, they may make a completely different decision. And so, that's the reason I have. My cynicism around this is that when we're trying and predicting a thing to do with human beings, there are so many different variables involved, it becomes really challenging to be very precise in those predictions. So that's the source of some of my cynicism there.

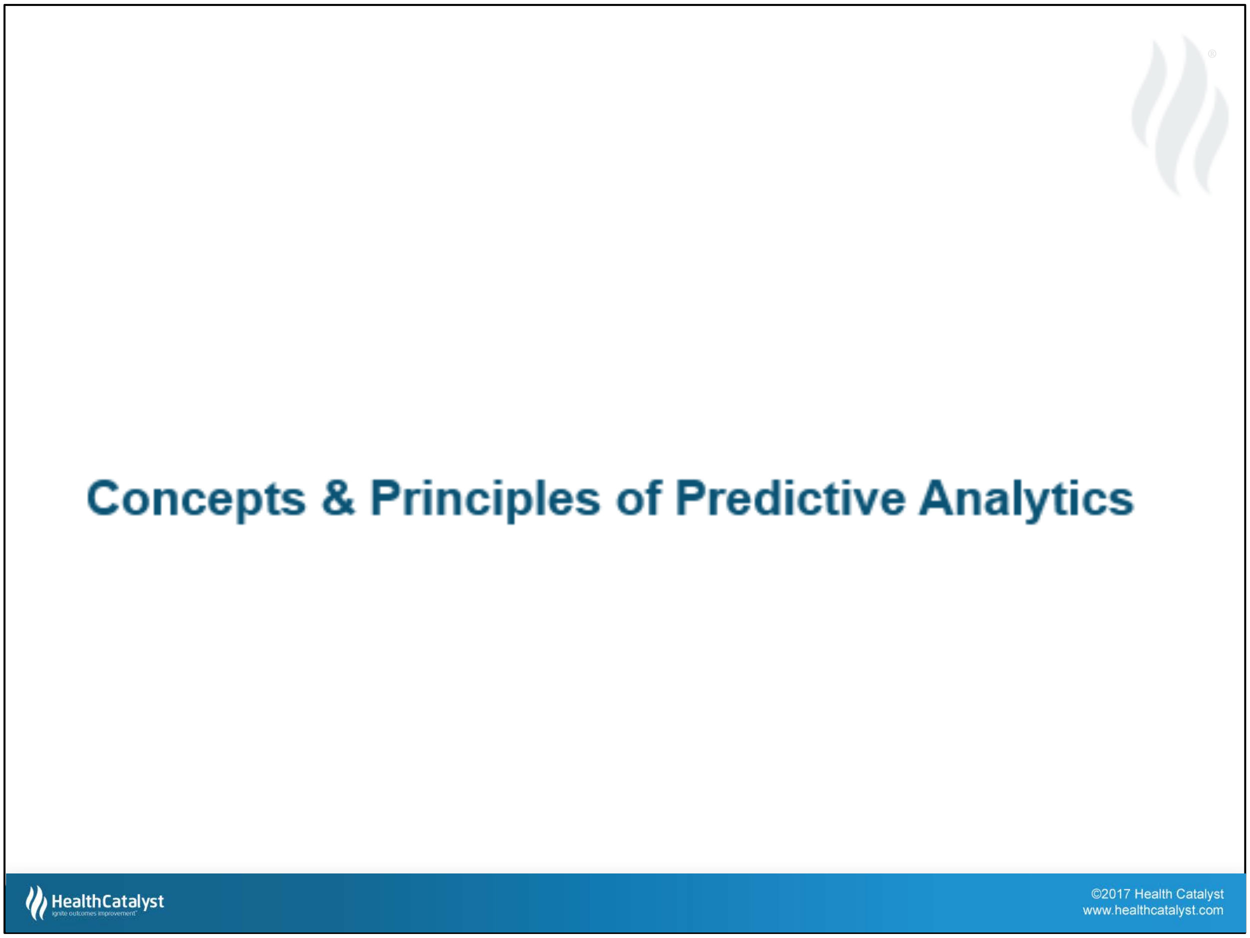


Healthcare Analytics Adoption Model [11:53]

This Healthcare Analytics Adoption Model is something that we developed over the last few years to help guide organizations through their progression towards very mature and sophisticated analytics. And you'll notice on this that we don't mention predictive analytics until level 7, almost at the top. And one of the reasons is because of this human nature problem that we have combined with the lack of adequate data, or I would say precise data in healthcare. So when you combine kind of a poor data environment with a human-centric environment, the value for predictive analytics is hard to achieve.

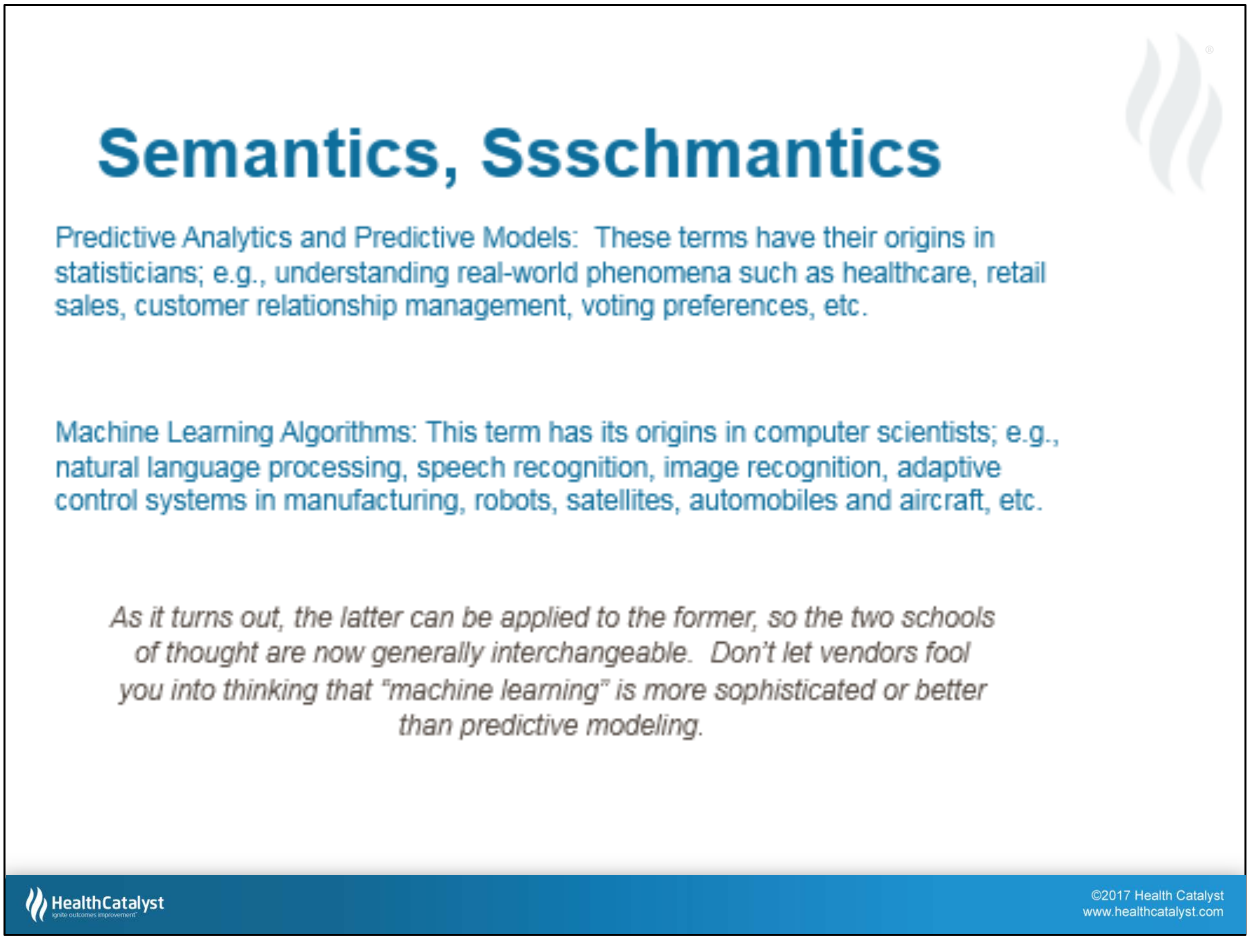
In the meantime, there are a lot of things that we could be doing to improve our reporting and analytics in the industry. They are a lot easier, a lot more achievable and a lot more repeatable. In particular, you know, I focused on level 5 here, we know that realistically there's at least 25% waste in every healthcare organization. There's probably more like 30% to 50% waste in some cases, depending on who you believe, with as much as $300 billion per year in wasted unnecessary care, sometimes even harmful care. And that doesn't require any predictive algorithms really to understand and start eliminating that waste. And if we did that, we would single-handedly solve the financial crisis that were in the US, if we just focus on level 5 analytics, and it's totally within our grasp to do that.

So as you plot your analytics strategy, I would encourage everyone to most definitely put predictive analytics on your roadmap but also focus on those things that we know we can do right now. They're not that complicated and they would have a big impact on both the quality, as well as the cost of care we deliver.



Concepts & Principles of Predictive Analytics [13:53]

So let's talk about some of the concepts and principles of predictive analytics in general.



Semantics, Ssschmantics [13:58]

Semantics, Ssschmantics. There are a lot of different terms being tossed around right now in the press. Predictive analytics and predictive models, those terms tend to have their origins from statisticians, trying to understand real-world phenomena such as healthcare, retail sales, customer relationship management, voting preferences, and that kind of thing.

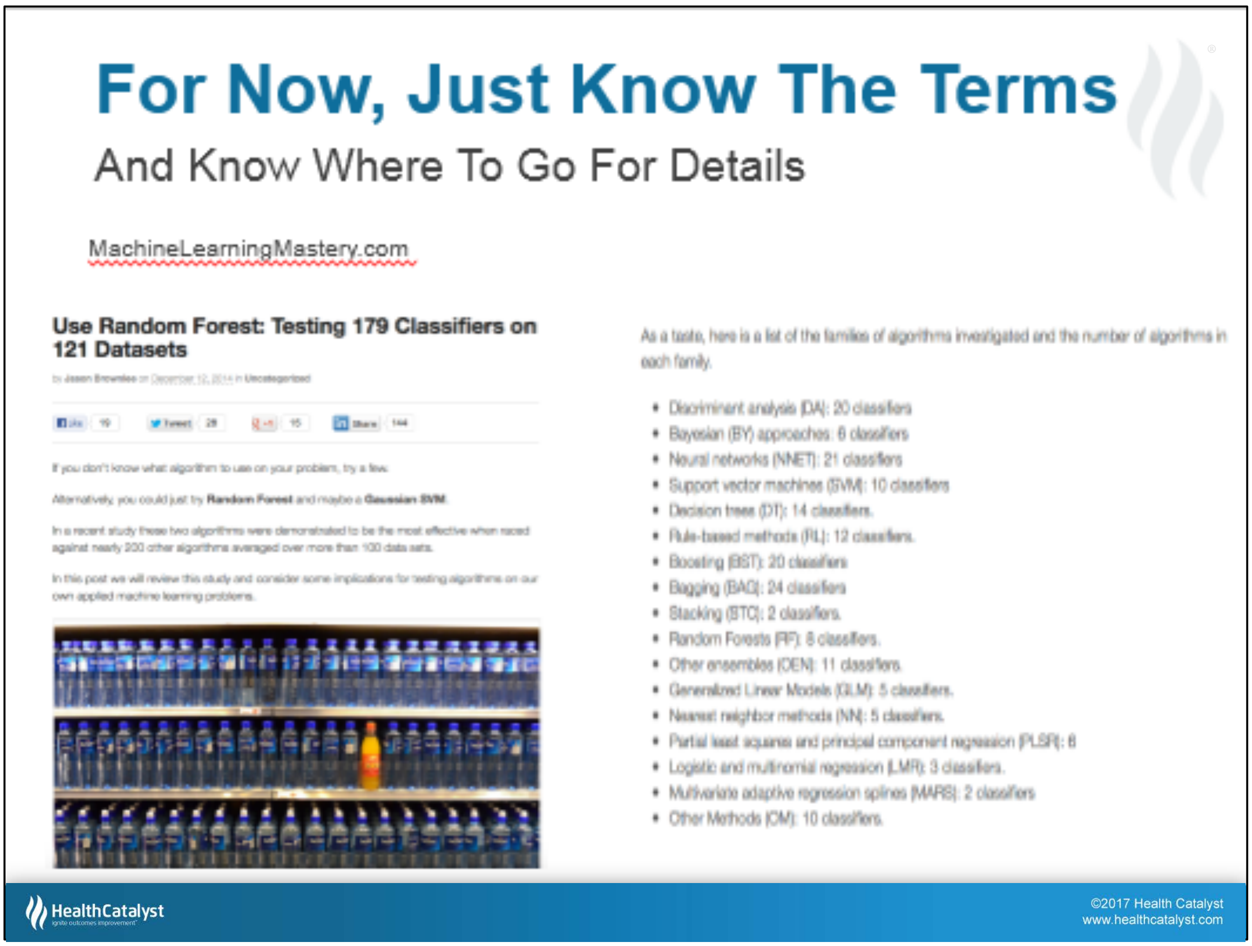
The other discussion and the other term that's becoming very popular is machine learning and machine learning algorithms, and this term has its origins in computer scientists. And this is, actually the world that I grew up in in the military was more around this and it was natural language processing, speech recognition, picture extraction from images, adaptive control systems in manufacturing robots, satellites, automobiles, and that kind of thing.

For Now, Just Know The Terms
And Know Where to Go for Details [14:49]

Well as it turns out, we found that the latter could be applied to the former. So now the two schools of thought are generally interchange one and I wouldn't – you know, one of the things I see vendors doing and some authors doing right now is they're talking about machine learning as if it's more sophisticated or it's better or something than predictive modeling. But the reality is they're both planning together and you have to consider both. And so don't be fooled by any of that. The two of them are becoming interchangeable and morphing together.

One of the things I was going to do in today's webinar was to give you kind of the pros and cons of the different techniques for predictive analytics. But I tell you, the more that I dug into that, the more I realize what a big undertaking that would be. So I've decided not to do that today, maybe leave it for another webinar where we just focus on a discussion of pros and cons.

For today's purposes and for most of the audience, the distribution of folks that I see in the audience, I would just encourage you for now be familiar with these terms and know where to go for the details. And again, the predictive analytics section on Wikipedia is awesome. I love the book that I've shown here from O'Reilly *'Doing Data Science'*. It's an extremely well-written book. It's very reachable from a broad audience. If you don't have a programming background, it's really good because they have a whole series of very simple-to-understand R programs that allow you to take concepts and see how that's reflected in software. The R, for those of you who don't know, is a very common open source tool for using predictive analytics and machine learning. So, just be familiar with these terms and know where to go for more details, and maybe we'll have a follow-up webinar that digs into these pros and cons of each of these different options a little deeper.



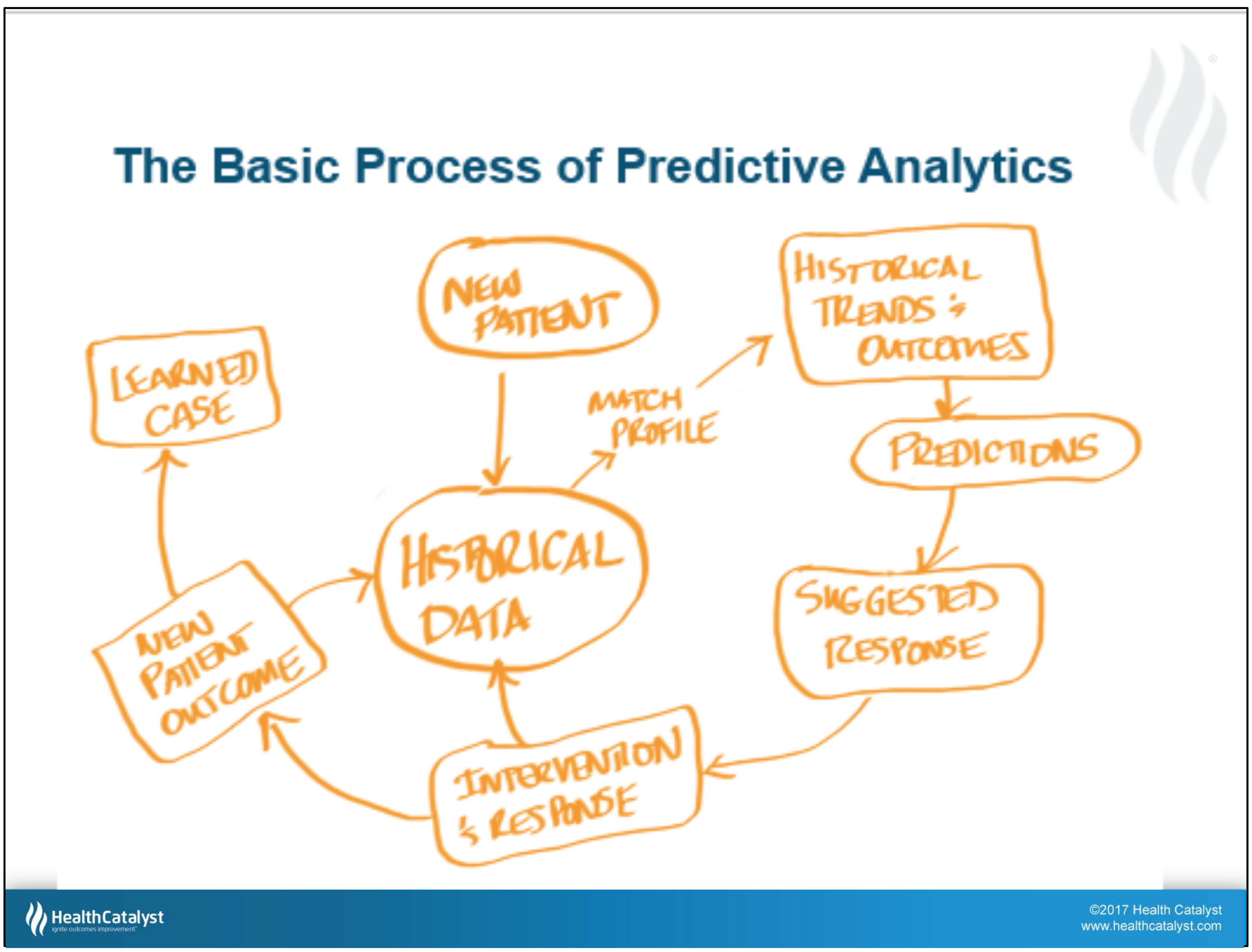
For Now, Just Know The Terms
And Know Where to Go for Details [16:56]

Another source that David Crockett shared with me just this morning actually is a site called MachineLearningMastery.com. And coincidentally they posted an article, I believe it was this week, earlier this week, on this very issue of how do you decide which is the hundreds of models that are available to us now, how do you decide which one of those to use. And this article is just awesome in its scope because they went through, oh my gosh, how many, 179 different classifiers and data sets, 121 data sets, to try to determine in general are there best machine learning algorithms that tend to hold up no matter what the data or use case. And in this case, they are concluding that Random Forest and this Gaussian Support Vector Machine

(SVM) are the two that tend to hold up the most no matter what the use case or the data content.

On the right side of this slide, you see all of the different families of algorithms. So these are just families of algorithms. And underneath those families and types of algorithms are dozens of classifiers that you can use. And this is where my knowledge and my expertise start to decay rapidly, at this level of detail. This is where people like David Crockett and Carl Anderson in our team, they stay up-to-date with this and they invest constantly in keeping up-to-date with this.

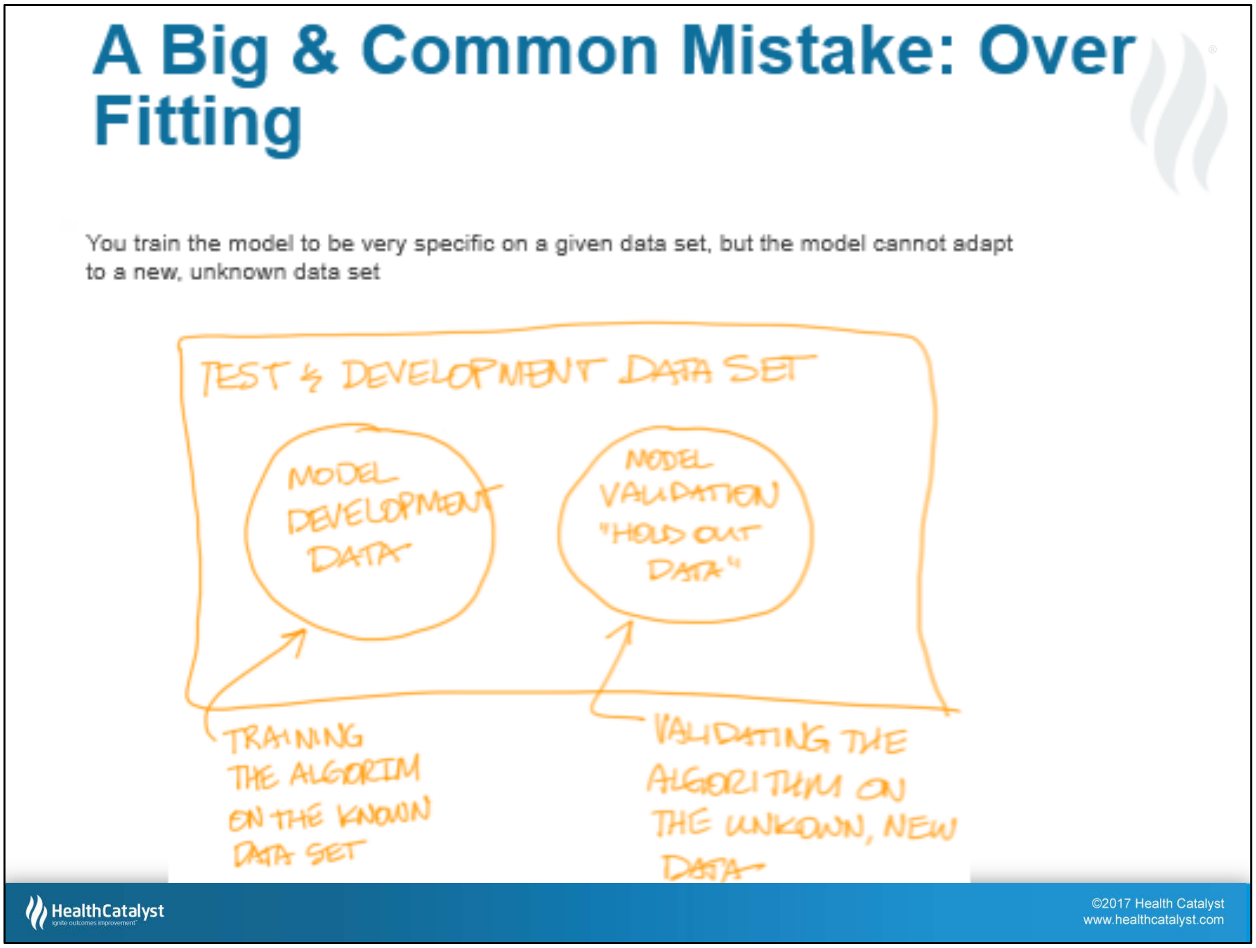
And as I look forward, you know, one of the reasons I'm a little cynical about predictive analytics in healthcare is the level of expertise required to keep up with this and to optimize the technology is very very deep. And I just don't know that we have the ability to spin up those resources fast enough to be accurate and precise without being dangerous with these predictions. So the skills area is one of my concerns going forward frankly. If you have any kids that are getting ready to go to college, this would certainly be a great area for them to study.



The Basic Process of Predictive Analytics [19:13]

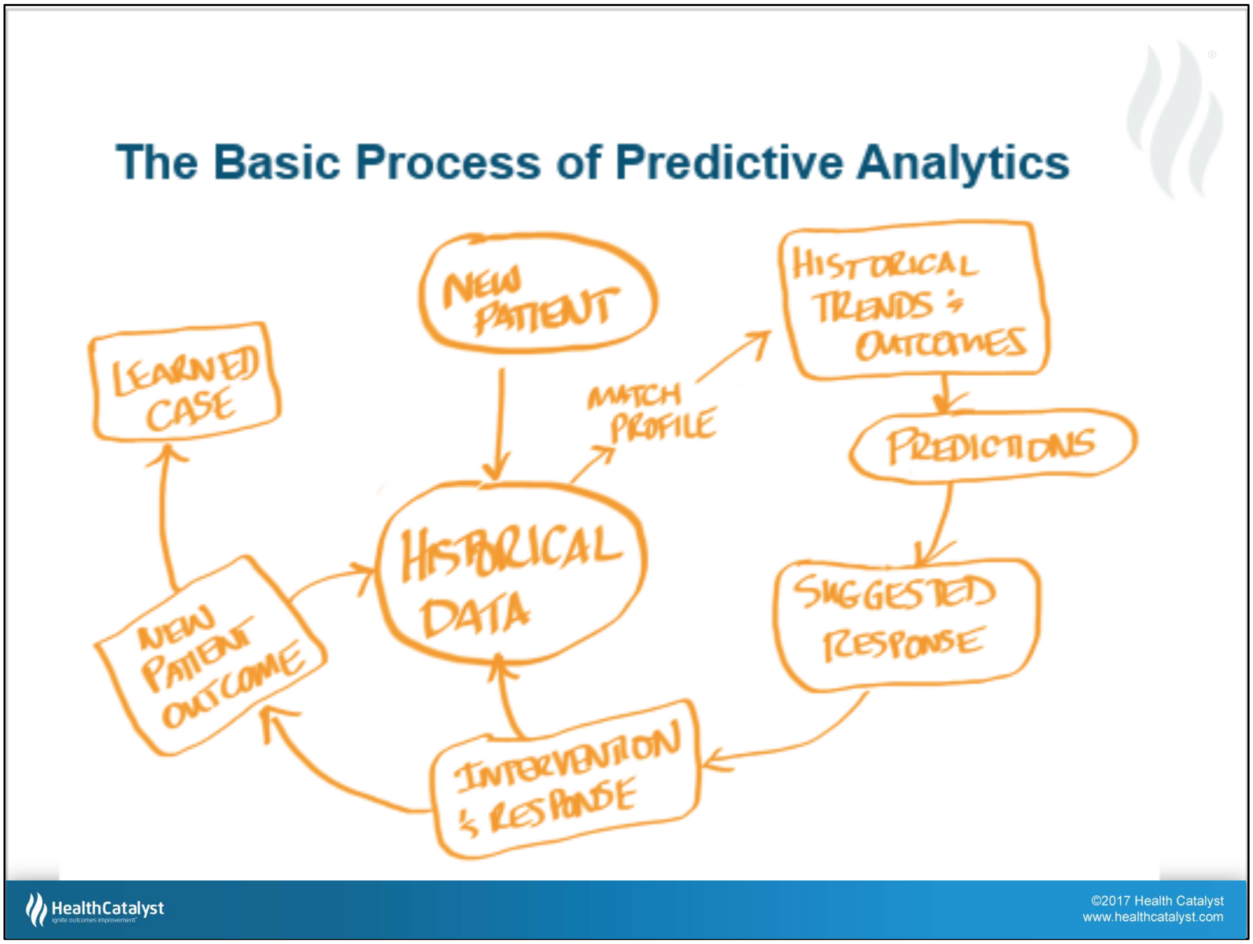
So this diagram summarizes just the basic process of predictive analytics no matter what algorithm or approach that you are using. This pretty much holds up no matter what, and that is, you know, starting in the middle here, a new patient arrives with their data profile accompanying them, you compare that to your historical data, match it to the profile, and the historical trends and outcomes associated with past patients of a similar type, and this is where the black fox of algorithms and machine learning exist, is in that predictions box right there. This box, suggested response, is actually threading upon prescriptive analytics. So it's one thing to predict an event for an outcome or a trend, but it's another to actually suggest a response to that prediction and this is where you start developing an intervention. And one of the key messages that I like to leave behind today is that the prediction without a strategy for intervention is useless. So as you're setting up your predictive analytics strategy, you have to think about what if we are successful and accurate with our predictions, what are we going to do with that knowledge? What are we going to do organizationally to intervene? I've got a good example, thanks to our friends in Allina, about how they developed an intervention strategy to reduce their readmissions.

So, once you've decided, and hopefully in a perfect world you have some sense of what your response will be and prescriptive analytics comes to play there, then you're going to form your intervention and response and you're going to fold that all back into the historical data. This is what we did, these are the medications we ordered, for example, these are the procedures that we ordered to treat and respond and intervene to this new patient. You fold that back into your historical data, and then hopefully you're also tracking that new patient's outcomes. Now, this is a big problem right now because we're not doing this in healthcare. We are not tracking outcomes in the extent, and patient satisfaction scores don't count as outcomes. The outcomes that we have right now are pretty much based on this belief that if you don't come back in for treatment, you must be okay. So it's the absence of outcomes we infer into directly something positive, but we need to be better about being more positive and more direct about those measurements – because otherwise it leads at worse in open loop in this process of managing data and at least a fuzzy loop that's only inferred data, not directly measured data. And then all this becomes a learned case and you keep going through this cycle. And generally speaking, the more data that you have here in the center, the more likely and the more accurate your predictions and prescriptions are going to be for intervention. Now, volumes doesn't always equal quality and we'll talk a little bit more about that in a few minutes. But generally speaking, the more data, the better.



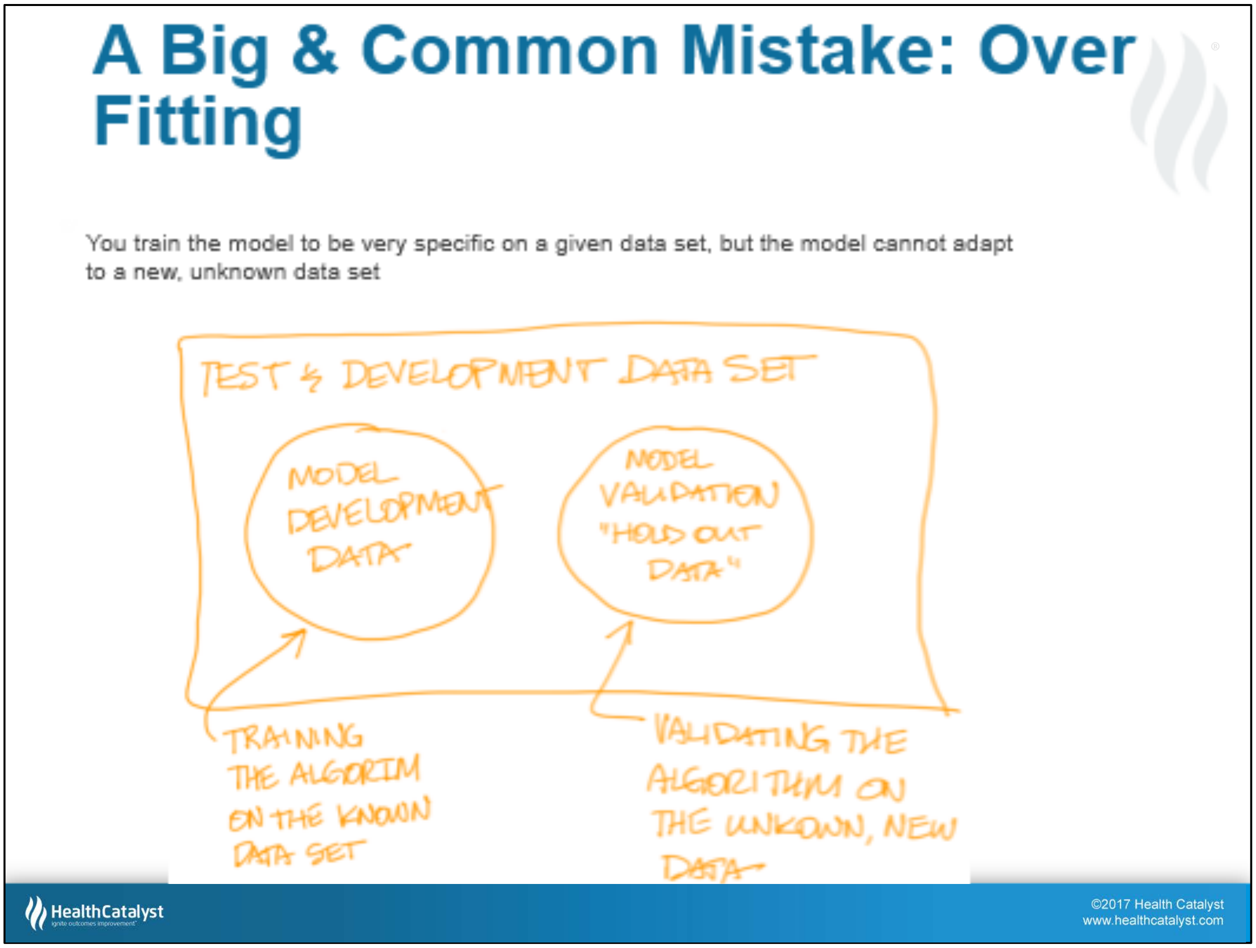
A Big & Common Mistake: Over Fitting [22:39]

So one of the big and common mistakes that I've seen happen, and I've made this mistake myself, is over fitting these predictive models so that they're very specifically trained around a given data set.



The Basic Process of Predictive Analytics [22:55]

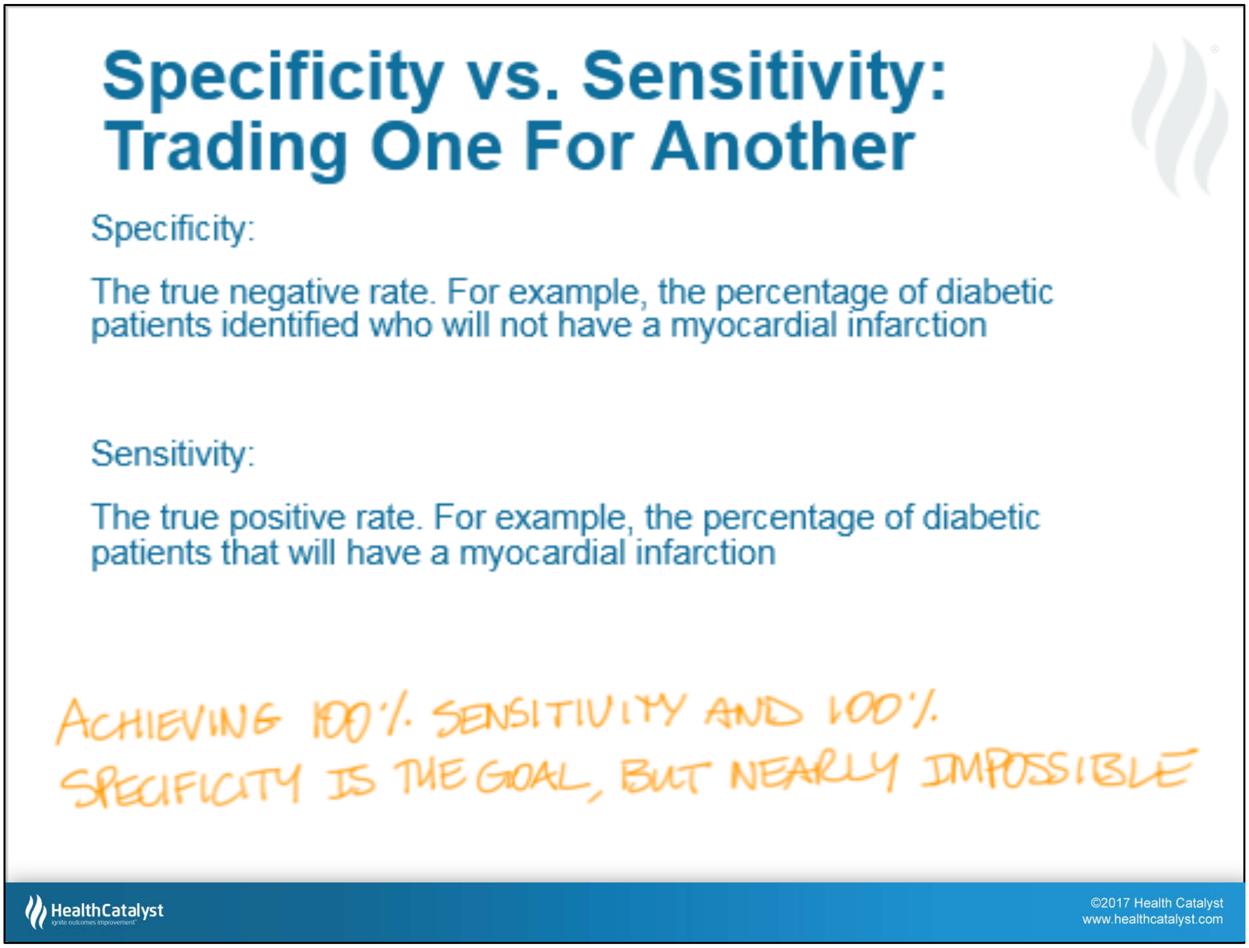
So let's go back for just a second to this historical data. So you train that prediction and the suggested response around that, a relatively narrow field of data within this central historical data set.



A Big & Common Mistake: Over Fitting [23:09]

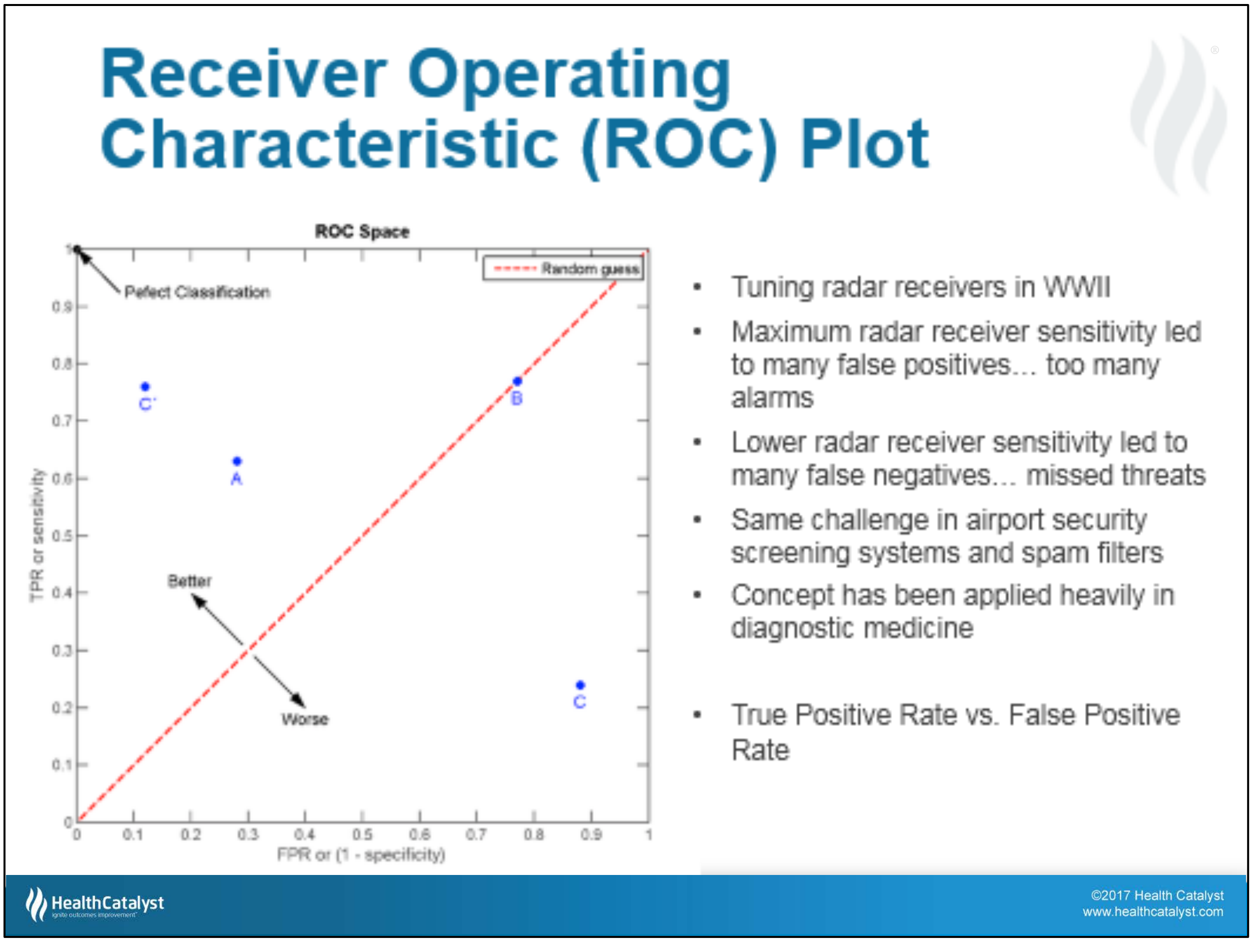
But the problem with that is those kind of models that are too specific, they become very difficult to generalize. So, this is actually an area where when you kind of look at the value that someone, like with my background who has a scale of 1 to 10 my technical skills are probably a 7 or something like that, the value that I've provided is in this setting up the development data set, understanding the data content, how that relates to these predictive models, and understanding data quality issues, that sort of thing. But the important point here is that you have to have both a model development data set and you also have to have a model validation "hold out" data set.

So this is the data that you're going to train your algorithms on and this is the data that you're going to validate those algorithms on. These are the new use cases. So really important not to suffer from over fitting. It's a very common mistake. I think generally it's becoming less of a problem. But for those, especially vendors who don't have much experience in this sort of thing and consumers who don't have a background in asking the right question that still think it's a risk that we need to be aware of.



Specificity vs. Sensitivity: Trading One for Another [24:32]

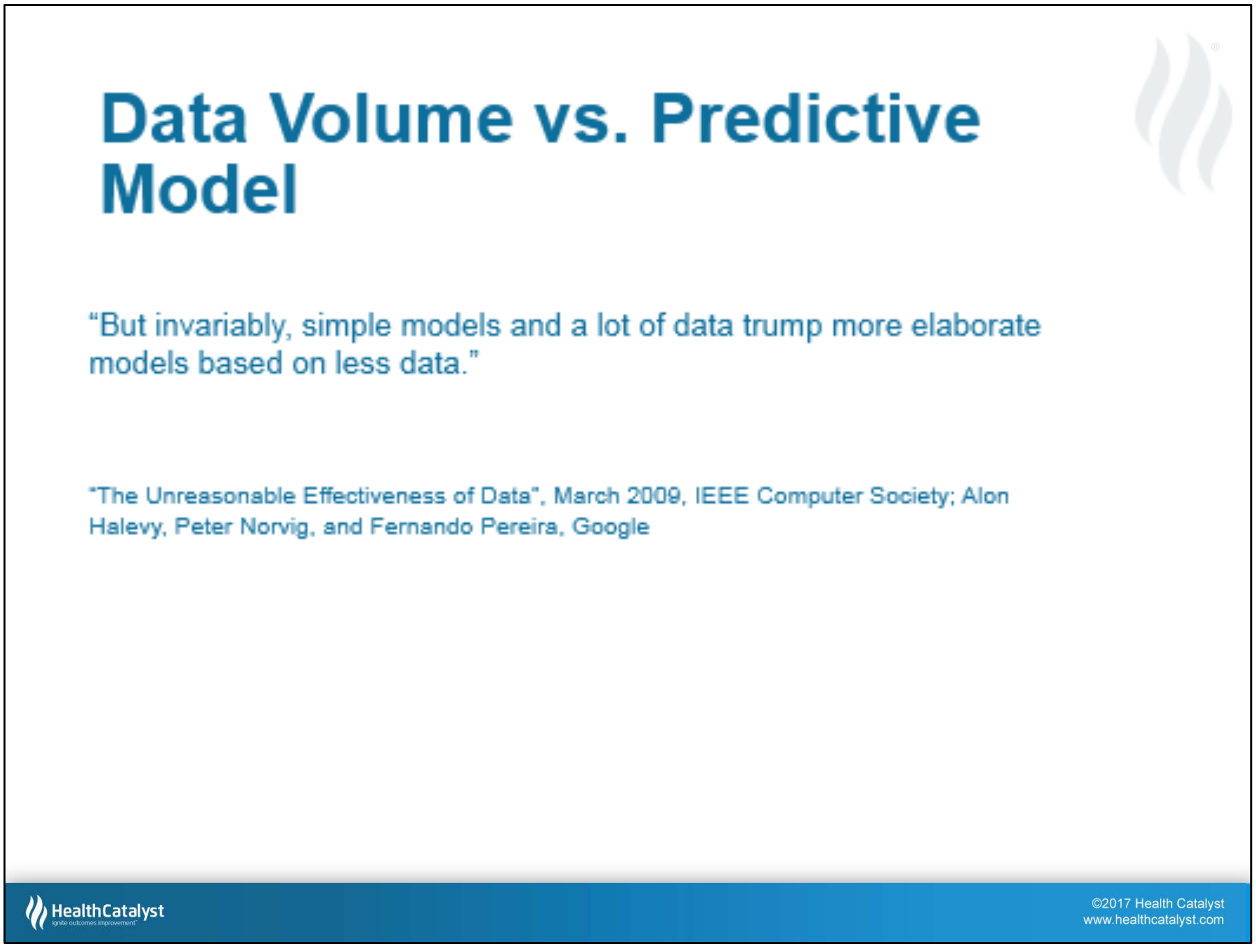
Okay. So related to this whole thing is this tradeoff between specificity and sensitivity. So, specificity in healthcare relates to the true negative rate. For example, the percentage of diabetic patients identified who will not have a myocardial infarction. The true negatives, right? What you're trying to avoid here are false positives and false negatives. Sensitivity is the true positive rate, and that is, for example, the percentage of diabetic patients that will have a myocardial infarction. And what we find is a little bit like the Heisenberg uncertainty principle here. It's a little hard to be perfect in either one of these without completely smearing the uncertainty in the other. Achieving 100% sensitivity and 100% specificity is challenging and this is where you'll want data scientists and data engineers that are capable of running the data through the algorithm, looking at the results, constantly adjusting the weights, constantly adjusting the data inputs to make this tradeoff between specificity and sensitivity. So a key question to ask of your analytics teams and your vendors is how are they handling this? Show me examples of where you tuned and you managed this and how you are going to do this in our setting.



Receiver Operating Characteristics (ROC) Plot [25:59]

Alright. So the evolution of this is kind of interesting. This is a ROC Plot. Some of you have probably seen this. It stands for Receiver Operating Characteristic and actually that term comes from its origins in World War II when they were trying to tune radar receivers. So, what they would find is that if they tune maximum sensitivity on the radar receivers, they're creating so many false positives, too many alarms. So then they would de-tune it, right? They'd lower the radar sensitivity and that led to false negatives, missed threats. So there's this constant tuning of these algorithms, you go back and forth, to try to try to find that balance. And the same challenge applies in airport security screening systems and spam filters, right? Too much specificity on your spam filters, too many things you're going to get through. Not enough specificity, too much of the other is going to allow too many things, they're not enough.

So, this concept has been applied quite heavily. You know, it's odd that it would come from radar receivers to medicine but you'll see this quite a lot. And what you want to ask of your vendors and your staff is how are we going to push our classification results into this part of the diagram up in this upper left area, the diagram, where the perfect classification is of its sensitivity of 1 and false positive rate of 0. Okay. So again, one of those concepts to ask about.

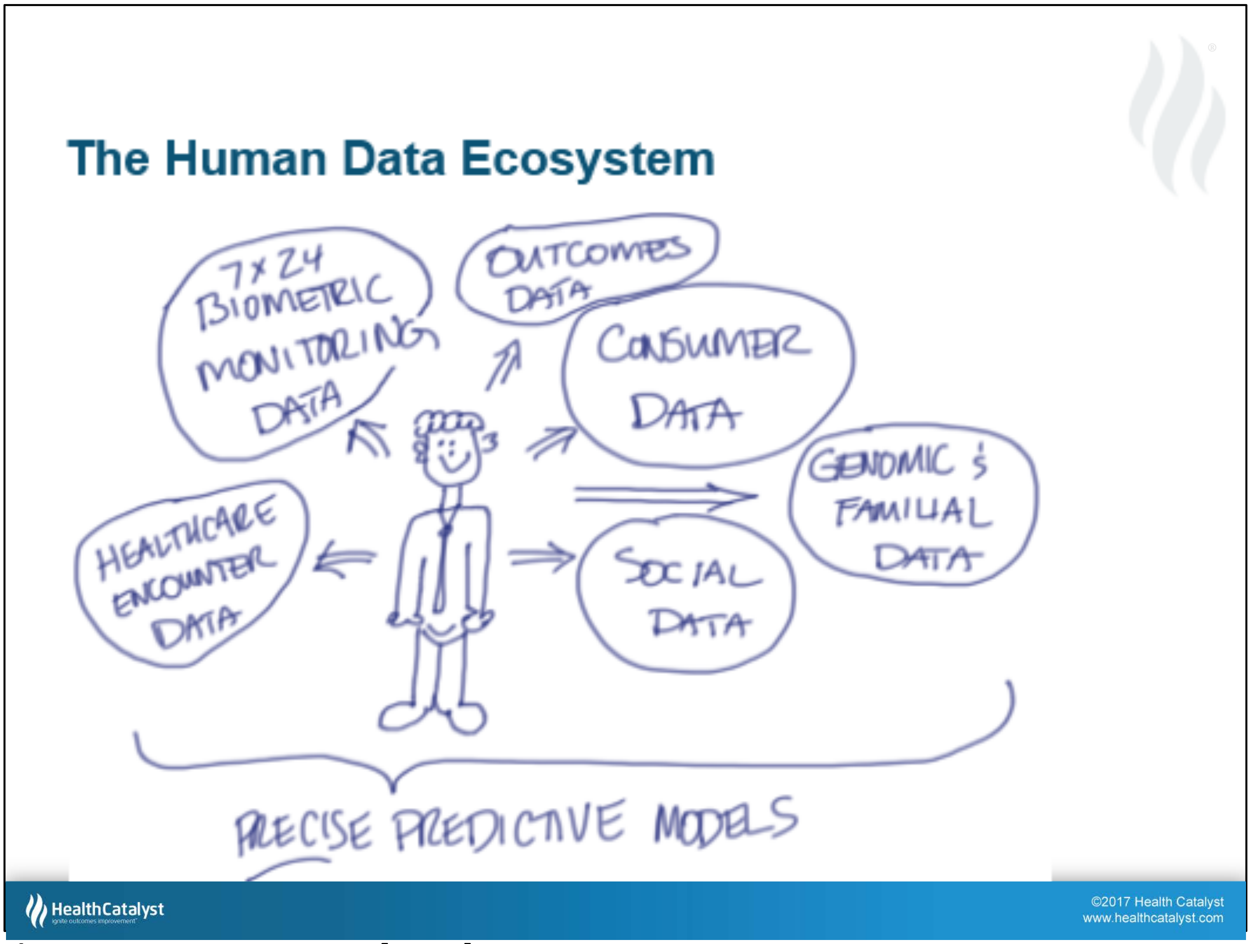


Data Volume vs. Predictive Model [27:40]

This is a quote from the folks at Google who published a really interesting paper in March of 2009, and they summed it up very nicely. They described this relationship between data volume and the predictive models. *"Invariably, simple models and a lot of data trump more elaborate models based on less data."*

So one of the things that troubles me is that I see some vendors and some authors writing now on these things called small data and medium data. And maybe you don't need as much data as everybody says you do but the reality is you need a lot of data and especially valuable. A data that's local to your organization and local to your reaching is the most important.

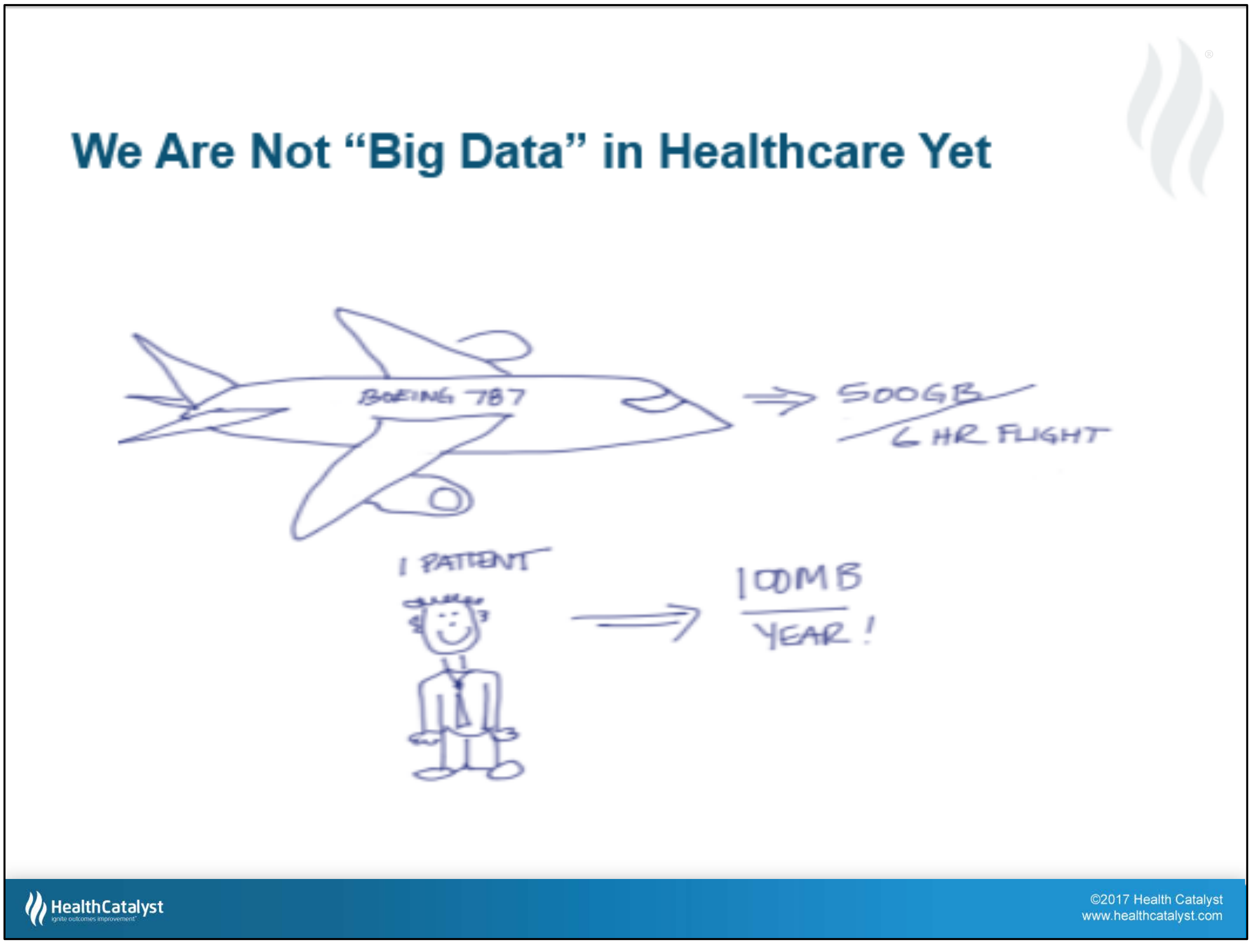
There is an inflection point on the curve, I think I have a diagram to that effect, where you can get into too much data that's not related to your environment. And it was kind of interesting to say in the mid-90's when we started seeing an overwhelming amount of data starting to flow in, and I can only imagine what the problem is now. But the predictive model started having a harder and harder time understanding that influx of data. And so, there is a tipping point where you can have too much data and actually start detracting from the act to receive your models.



The Human Data Ecosystem [29:11]

In this context, I'd like to mention that we tend to think of ourselves in healthcare as being big data, but the reality is we're not. There is a whole ecosystem of data that surrounds patients that's much broader than what we typically collect. Healthcare-encountered data is what we collect right now, but there's all sorts of opportunities outside of those encounters to collect more data to better understand this human being in the middle. We're starting to see some progress on 7 x 24 biometric monitoring data. I think the most important missing data in this whole thing is still outcomes data, activities of daily living, functional health status, SF-12 and SF-36, that kind of thing. I think an interesting area if I don't know if we'll be capable of doing it culturally but looking at consumer purchasing data, that Acxiom and Experian process, and trying to fold that into our risk model, I think has some opportunity for reducing the cost of care and improving quality of care. I just don't know if we're going to be prepared for it culturally. I would personally like to have the opportunity to enroll on a program like that if I knew it was going to result in lower cost and better for me and my family.

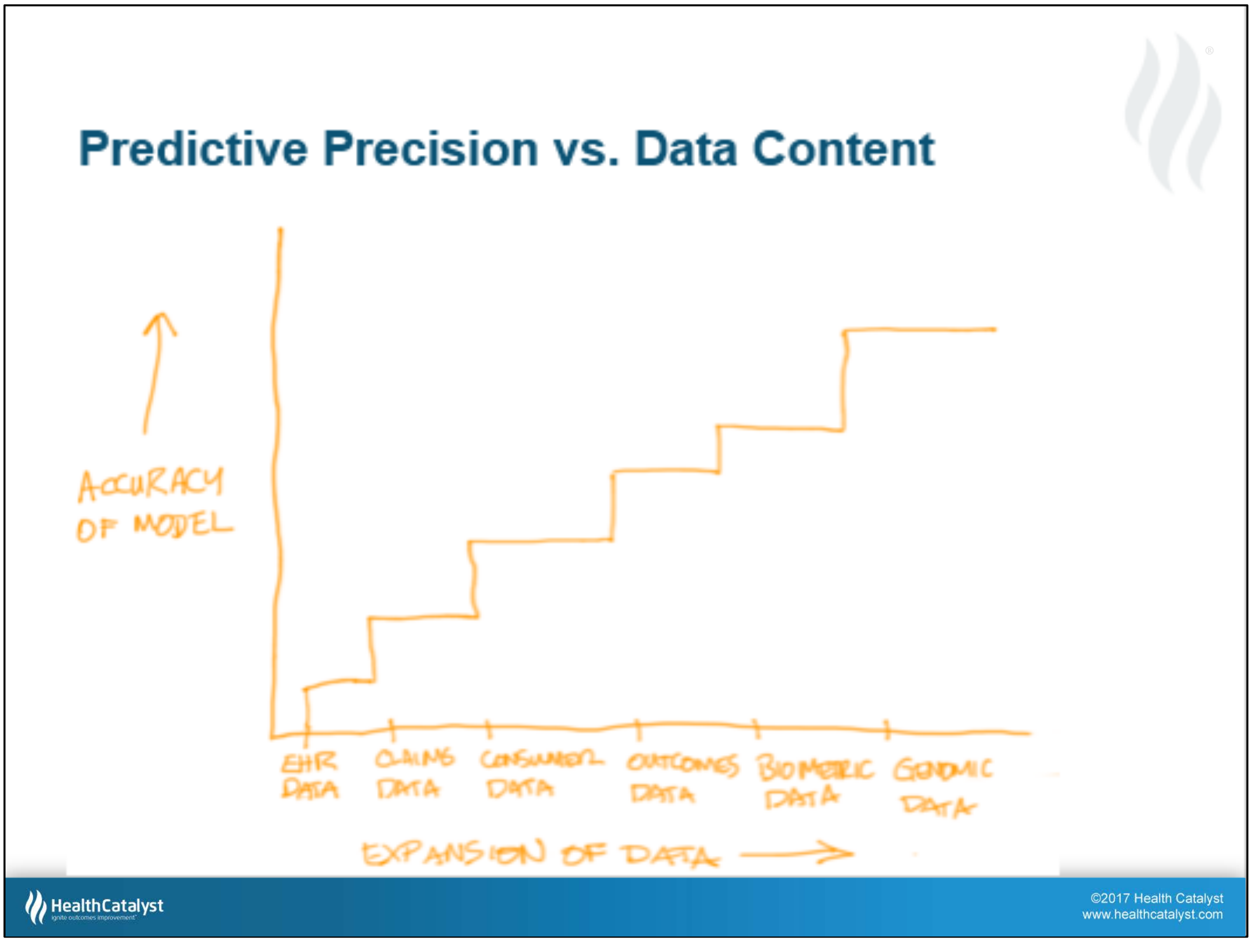
Genomics and familial data, in particular, they're going to have a huge impact on predictive analytics. We're still trying to get our arms around that but getting better and better all the time. And then the last vestige is social data. I don't have as much hope for that but it's certainly a different kind of data. I watched my wife go through her pregnancy the last year or so and the social data that she shared online with other pregnant mothers was fascinating, and with just a few little adjustments to that interaction, that data could be easily minable for better insights into outcomes and the effectiveness of medications and protocols and care pathways and things. So I do think there is an opportunity with some adjustments to that social engagement platform for patients that will have a big impact on our precision in predictive models.



We Are Not "Big Data" in Healthcare Yet [31:30]

Along this beam, I would like to mention, and some of you have seen this before, you know, we're not a "big data" in healthcare yet. Boeing 787 collects 500 Gigabytes of data in a 6-hour flight. A Formula One race team will collect 2 Terabytes of data in a weekend of test runs before a race. And I can speak first hand from that bottom graphic, when I was trying to calculate storage requirements in our EMR, excluding image data, that's a different animal, but imaging data is not that minable for predictive algorithm. So that's one of the reasons I excluded it. But we only collect about 100 MB of patient data per year. And for someone like me, I can't even think of the last time I've been in to see my physician.

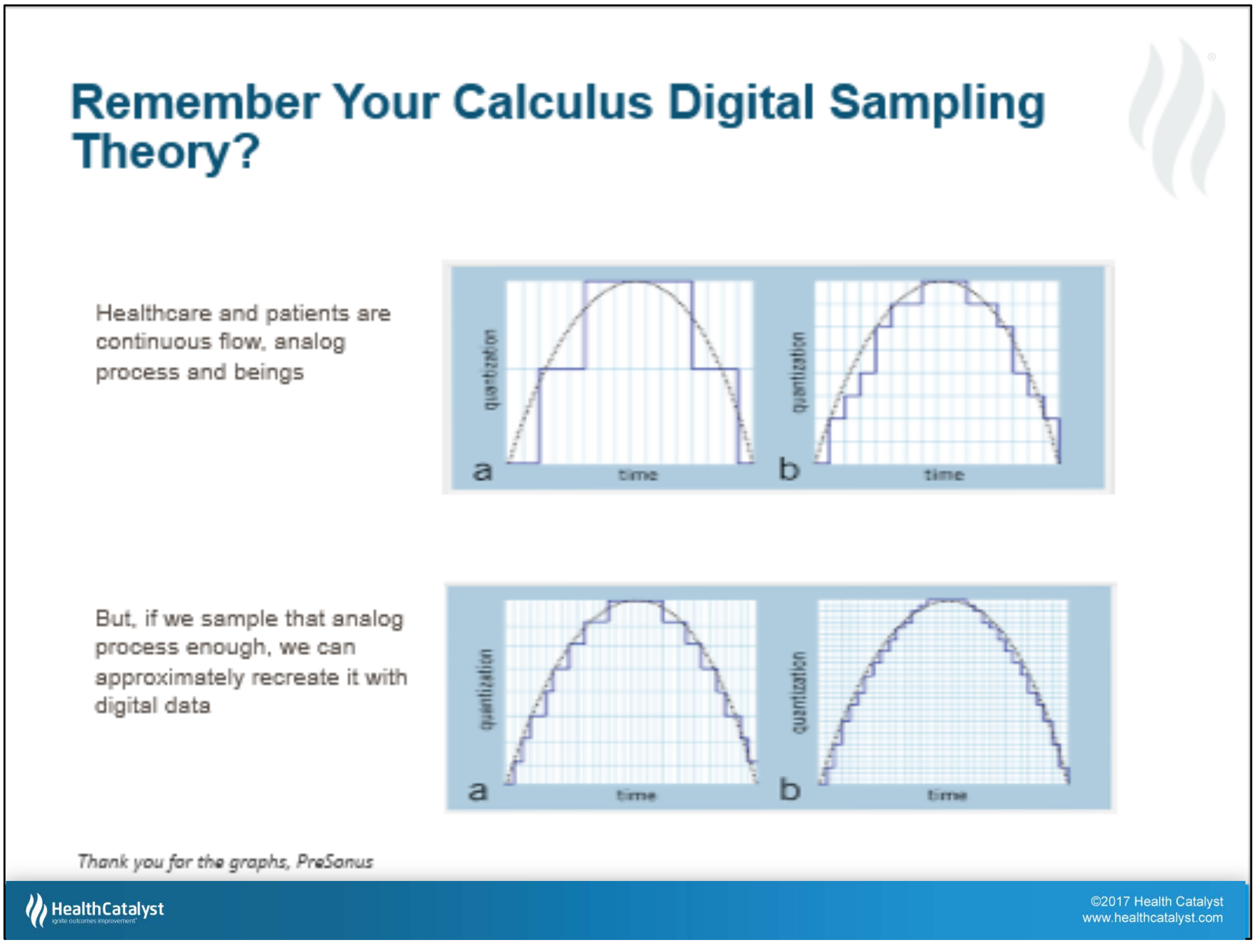
So, we're not sampling data at a rate that would make for better predictions, like a Boeing and other areas are. So that's another reason I'm a little bit cynical. We just don't have the data that we think we do.



Predictive Precision vs. Data Content [32:33]

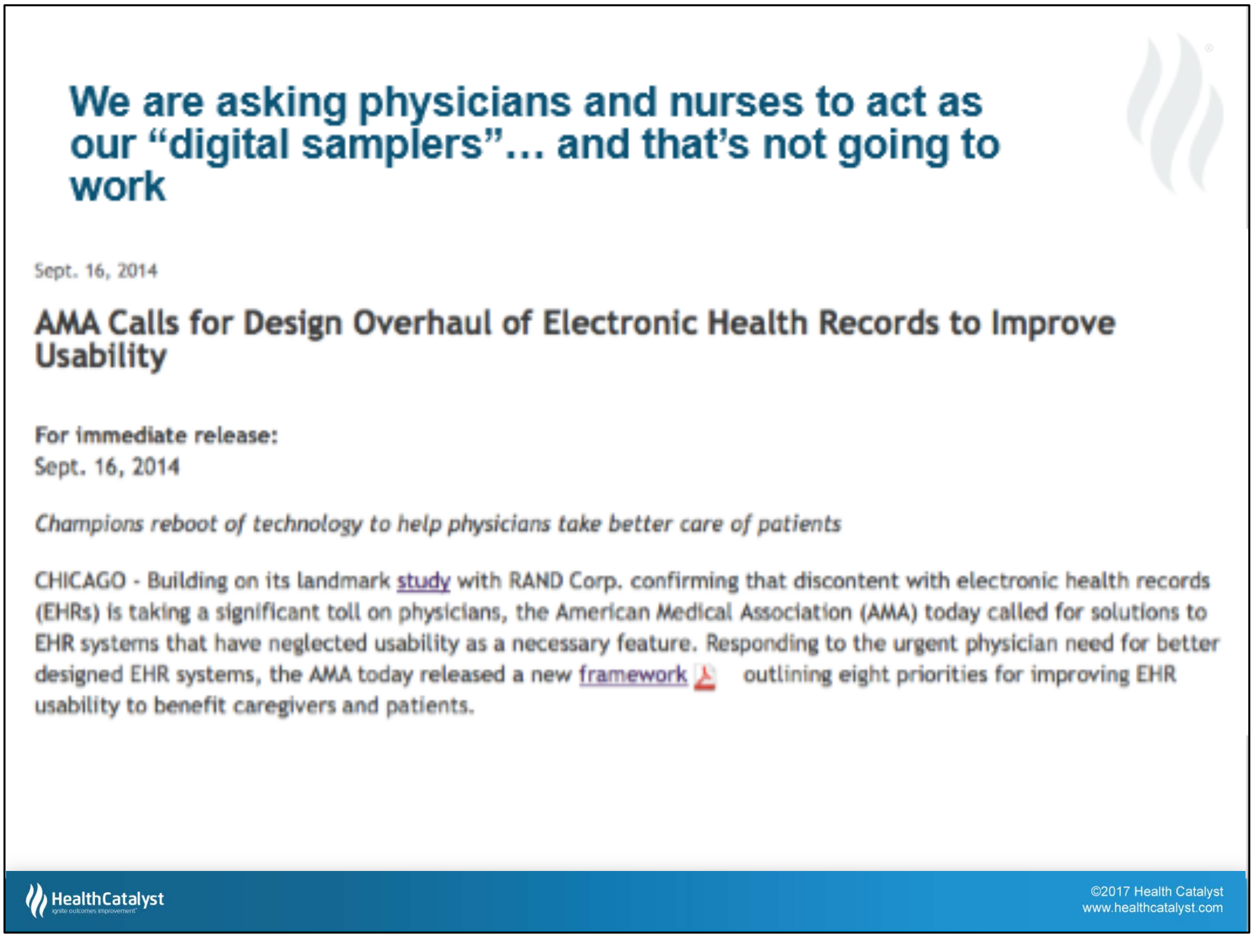
And this is a graphic and it shows that. So the accuracy of your predictive model tends to be related to the content that you have access to. And it will jump incrementally almost at a quantum fashion each time we add one of these sources of the data. So, EHR data might be the place that you start first and you'll get some value out of that in the accuracy of your predictive model. You have claims data. It's going to jump up at a quantum leap above that. Consumer purchasing data, outcomes data, biometric data, and so on. I don't even know what the rest of this looks like. But if you're a data governance function in a healthcare organization, you need to be thinking about how you aggregate these different sources of data. This is, in essence, the strategic acquisition for data roadmap. And if you have high hopes for the accuracy of a predictive analytics strategy, you'll think about the acquisition of this data as it relates to that accuracy. And if you think about, and I do believe this, organizations will be differentiated in the future around the effectiveness of their predictive models in terms of reducing clinical risks, improving clinical outcomes, and reducing financial risks.

So these are going to become very proprietary technical advantages to the early adopters, and those folks who have the deliberate planning place for acquiring data can make that even more differentiating.



Remember Your Calculus Digital Sampling Theory? [34:10]

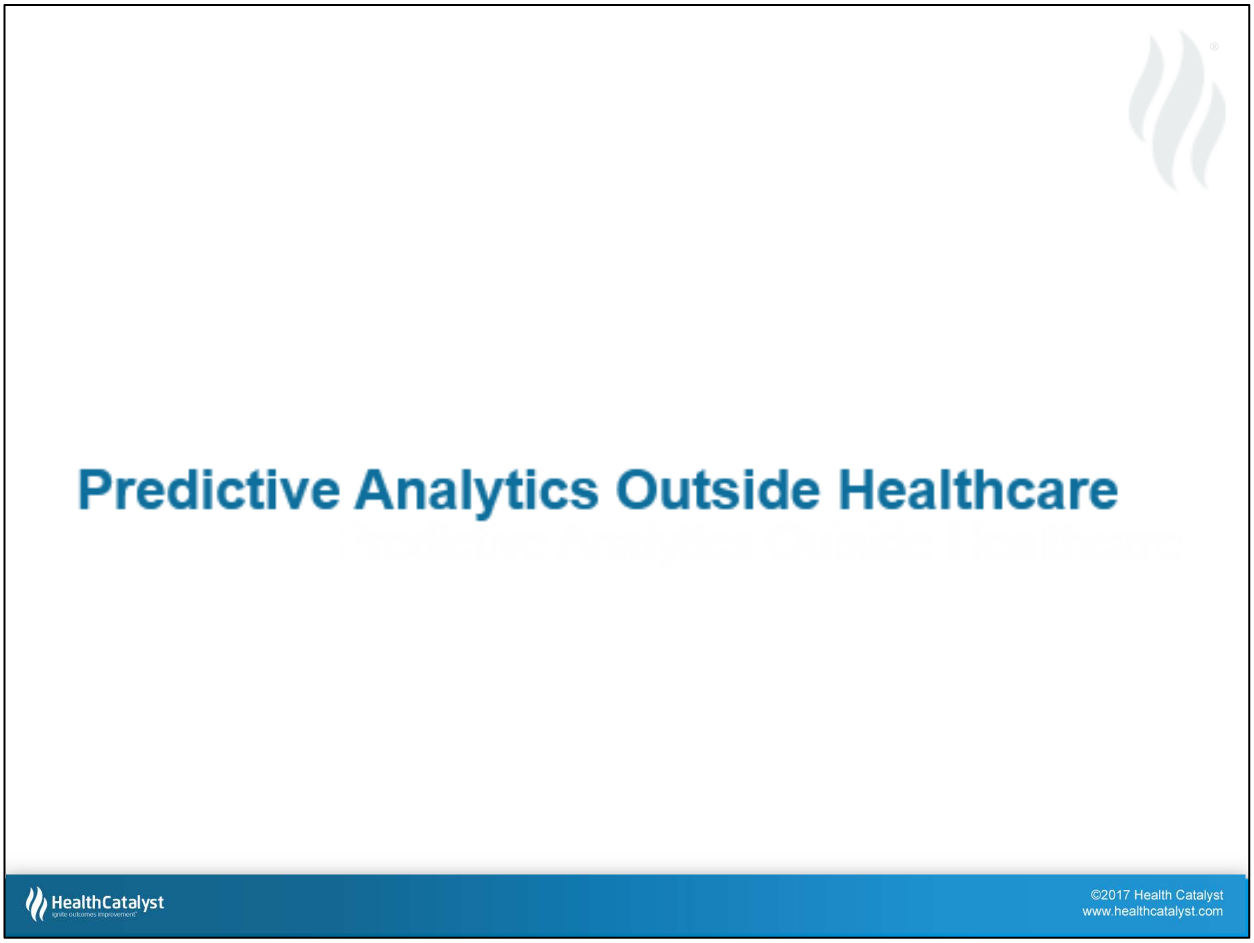
So just a quick comment here related to data content and that sort of thing. I'll go back to our old Calculus stage. You may remember, those of you that took Calculus Digital Sampling Theory, and you know, the basic premises here that healthcare and patients are continuous flow, analog process and beings. We don't exist in digital states. Now, some people would argue philosophically that we exist in quantum states at the sort of atomic level. But at the human level, we're analog creatures. And so, our lives are described by these smooth waveforms and what we're trying to do in healthcare through EMRs frankly is we're trying to sample that continuous existence of the human being with little snippets of data. And the notion is in digital sampling theory, if you sample that waveform often enough, your digital sampling starts to look more and more like the smooth analog waveform. Music is the classic example of this. In the old days, we had records with grooves. Those were a smooth analog waveform actually cut into the vinyl. Now, we have CDs and those CDs are just a digital sample of that analog voice and musical instrument represented in these samplings.



We are asking physicians and nurses to act as our "digital samplers"
…and that's not going to work [35:45]

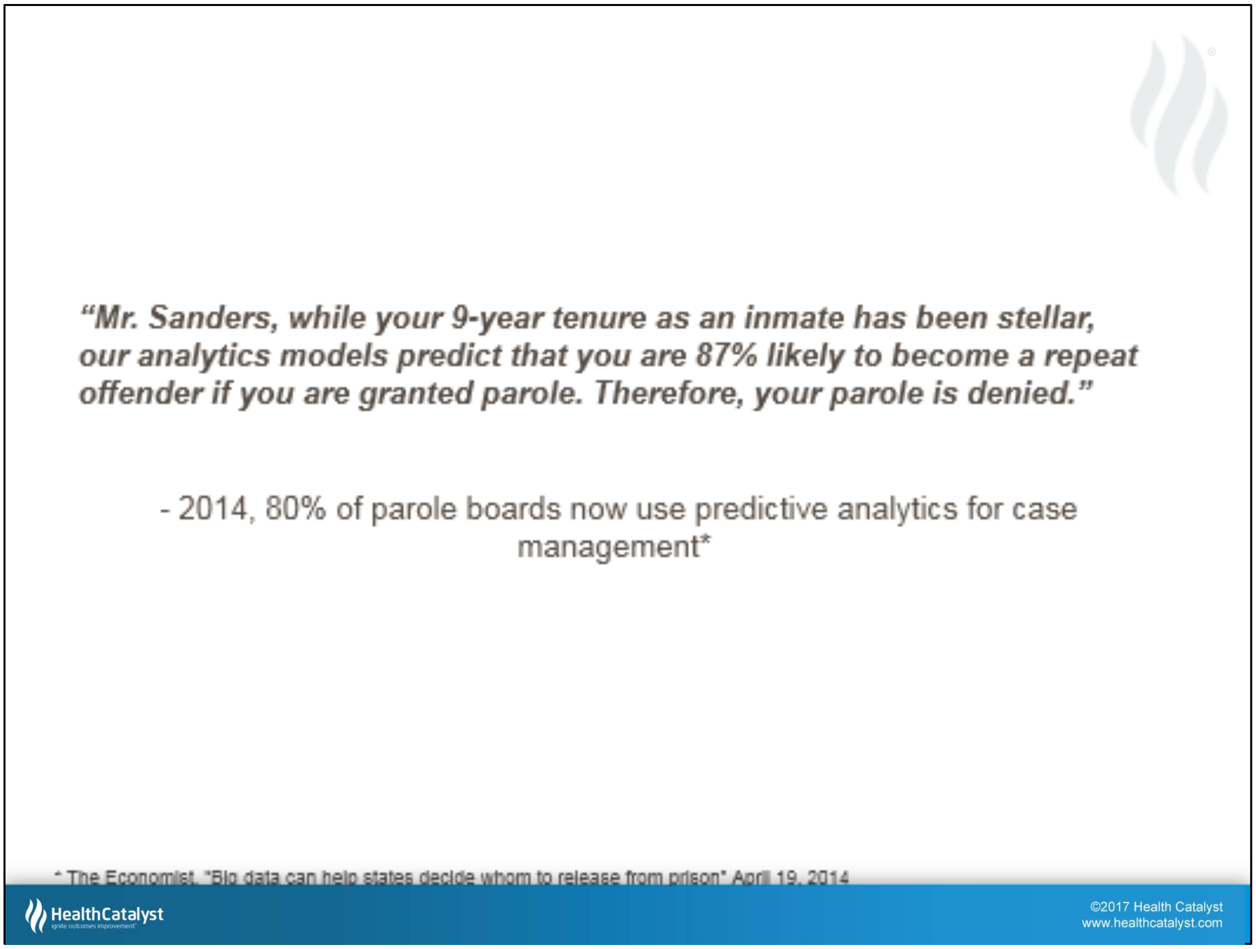
So one of the challenges that we have is we are asking physicians and nurses to act as our "digital samplers" and it's just not going to work. We can't expect them to keep continuing to click and click and click and collect more and more discrete data that represents those patient's experience.   We just can't do it. We have to figure out a way to reduce the number of clicks. And of course, the AMA published a report recently talking about how kind of poor the usability of EMRs is in this case.

So, we need to start thinking about how we're going to constrain the burden that electronic health records place on this digital sampling problem for nurses and physicians and start finding other ways to smooth out that analog waveform.



Predictive Analytics Outside Healthcare [36:29]

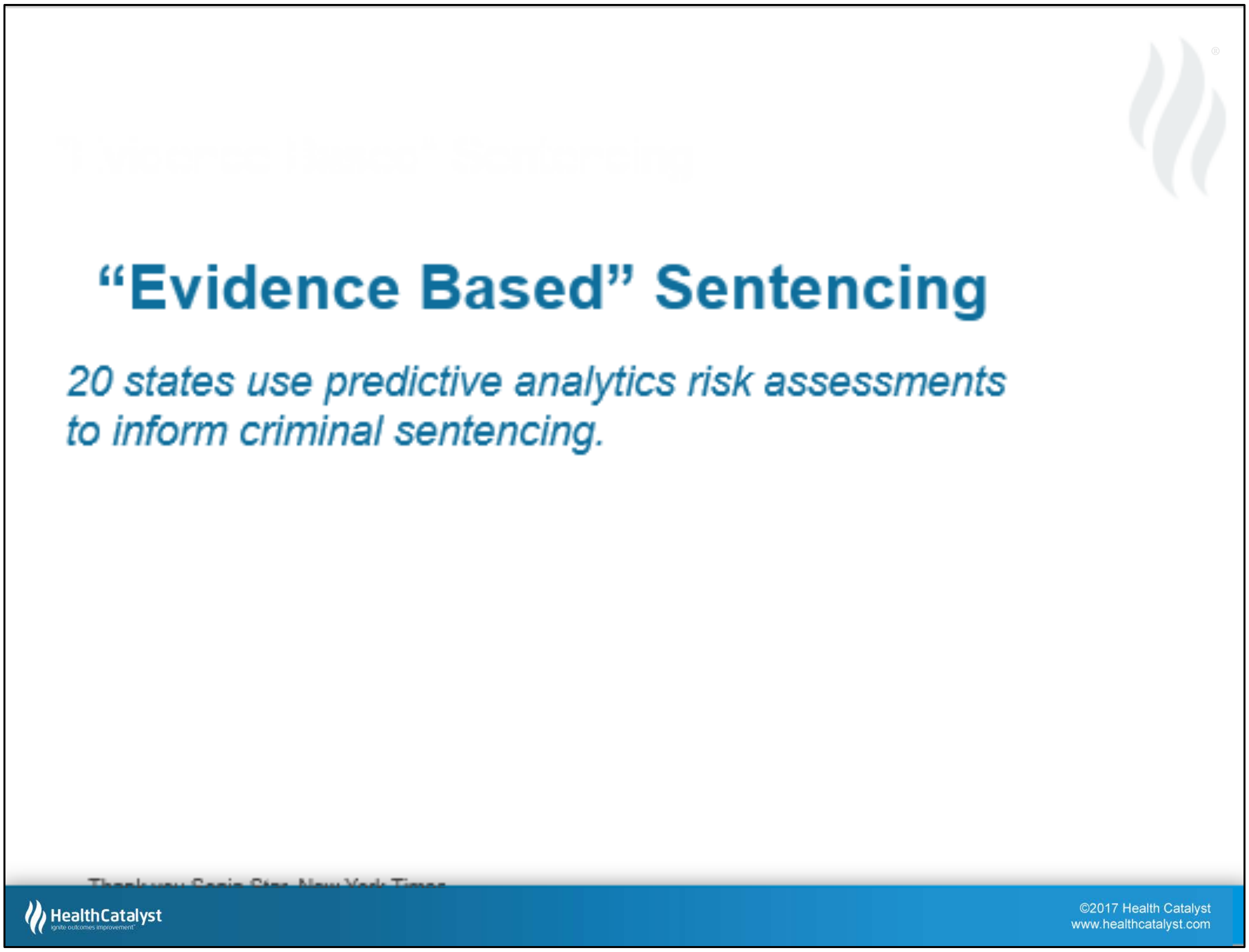
So let's talk a little bit about predictive analytics outside of healthcare.



[36:33]

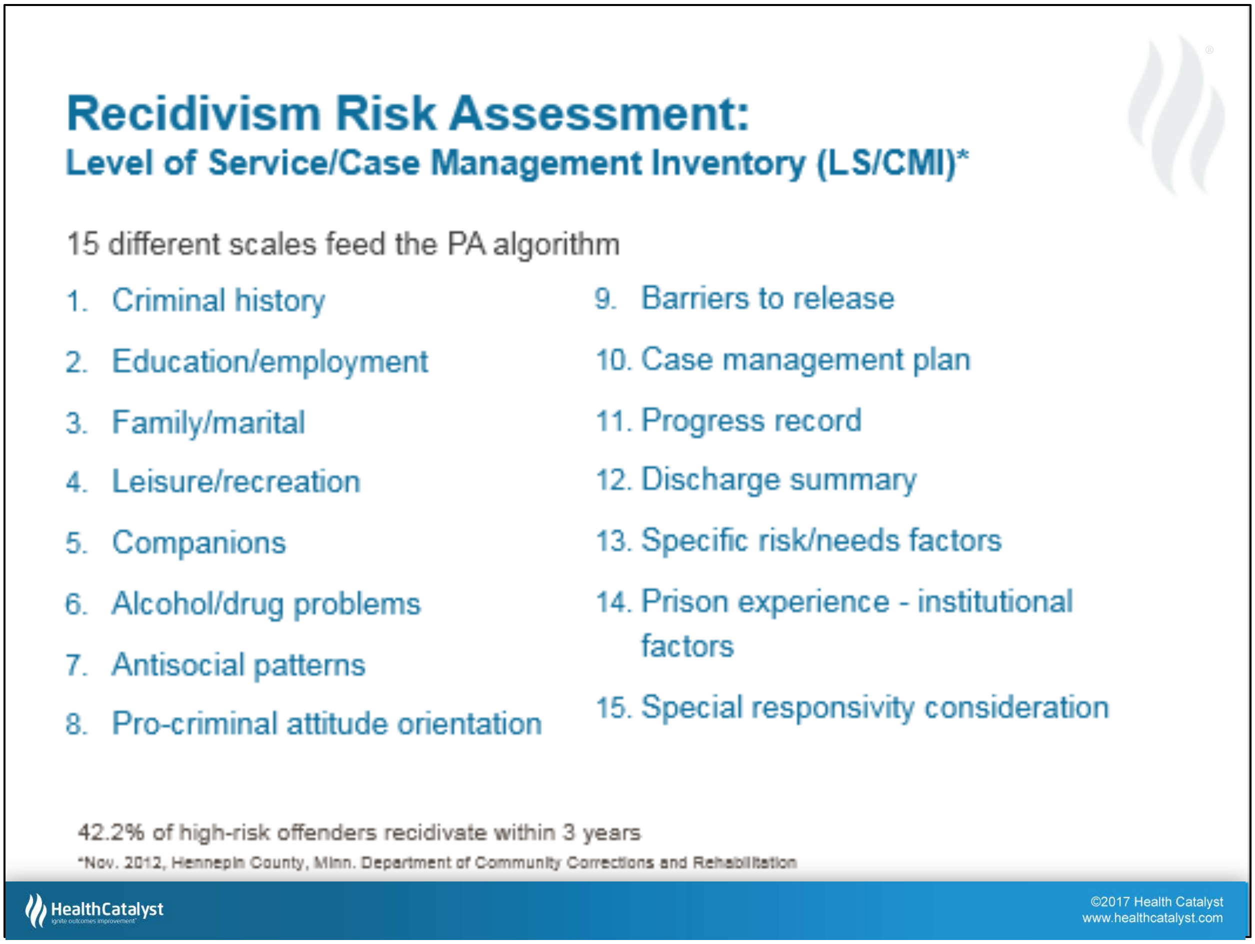
This is another kind of tongue-in-cheek sense of humor here from the autonomist. *"Mr. Sanders, while your 9-year tenure as an inmate has been stellar, our analytics models predict that you are 87% likely to become a repeat offender if you are granted parole. Therefore, your parole is denied."*

And so, it's kind of big brother issues, as it sounds, and/or willing as it sounds, it's happening right now. 80% of parole boards are now using predictive analytics for case management. And the reason I'm bringing this up friends is because I think it foreshadows the cultural challenges that we're going to face when we start using predictive models more aggressively in healthcare and you'll see a little bit more, I think, about what I'm saying here.



“Evidence Based” Sentencing [37:22]

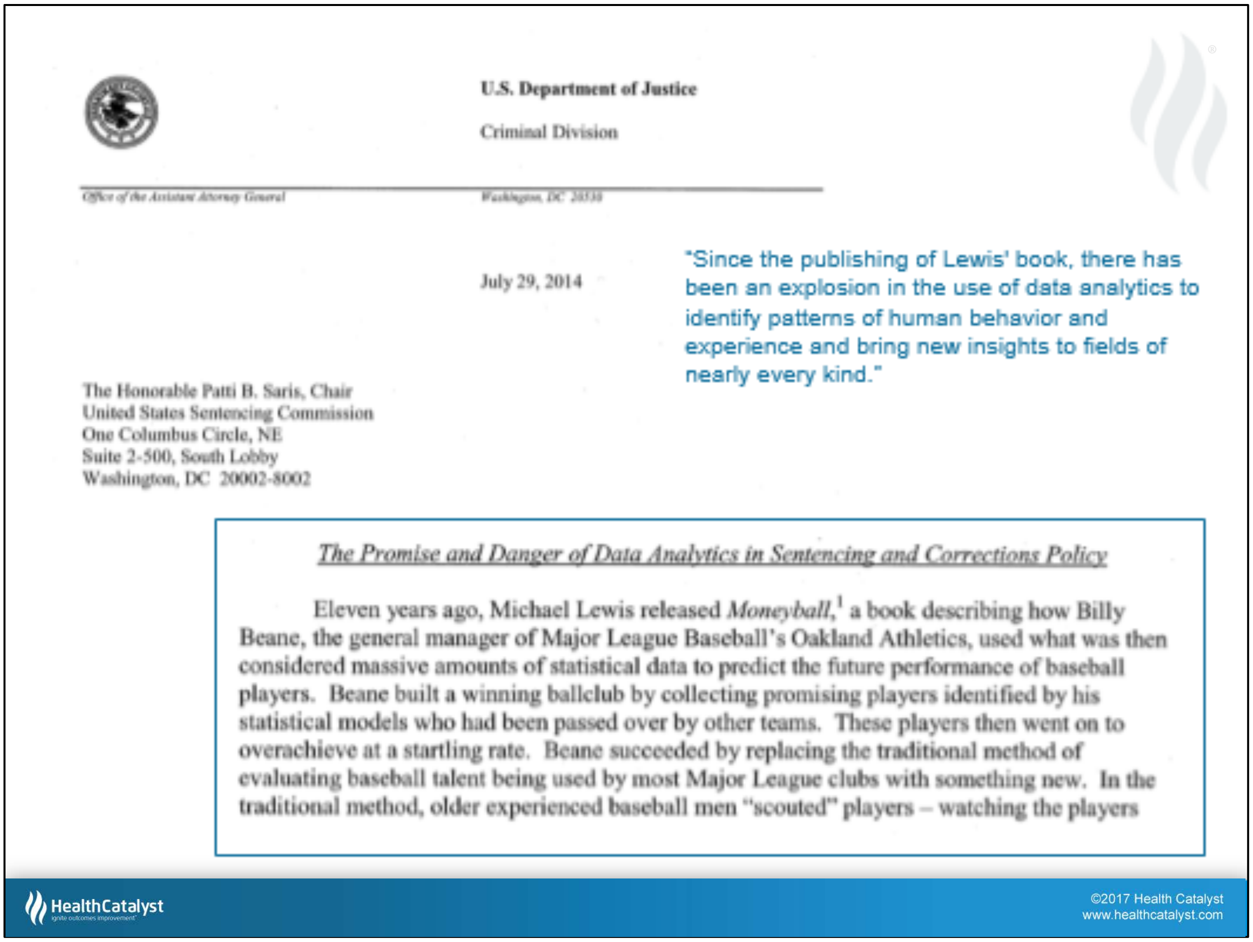
So, 20 states are now using “evidence-based” sentencing based around predictive analytics, trying to reduce recidivism. That’s their version of a readmission. Readmission in healthcare is recidivism in sentencing an imprisonment. So it’s already happening and it’s kind of happened quietly without a lot of attention.



Recidivism Risk Assessment:
Level of Service/Case Management Inventory (LS/CMI)* [37:42]

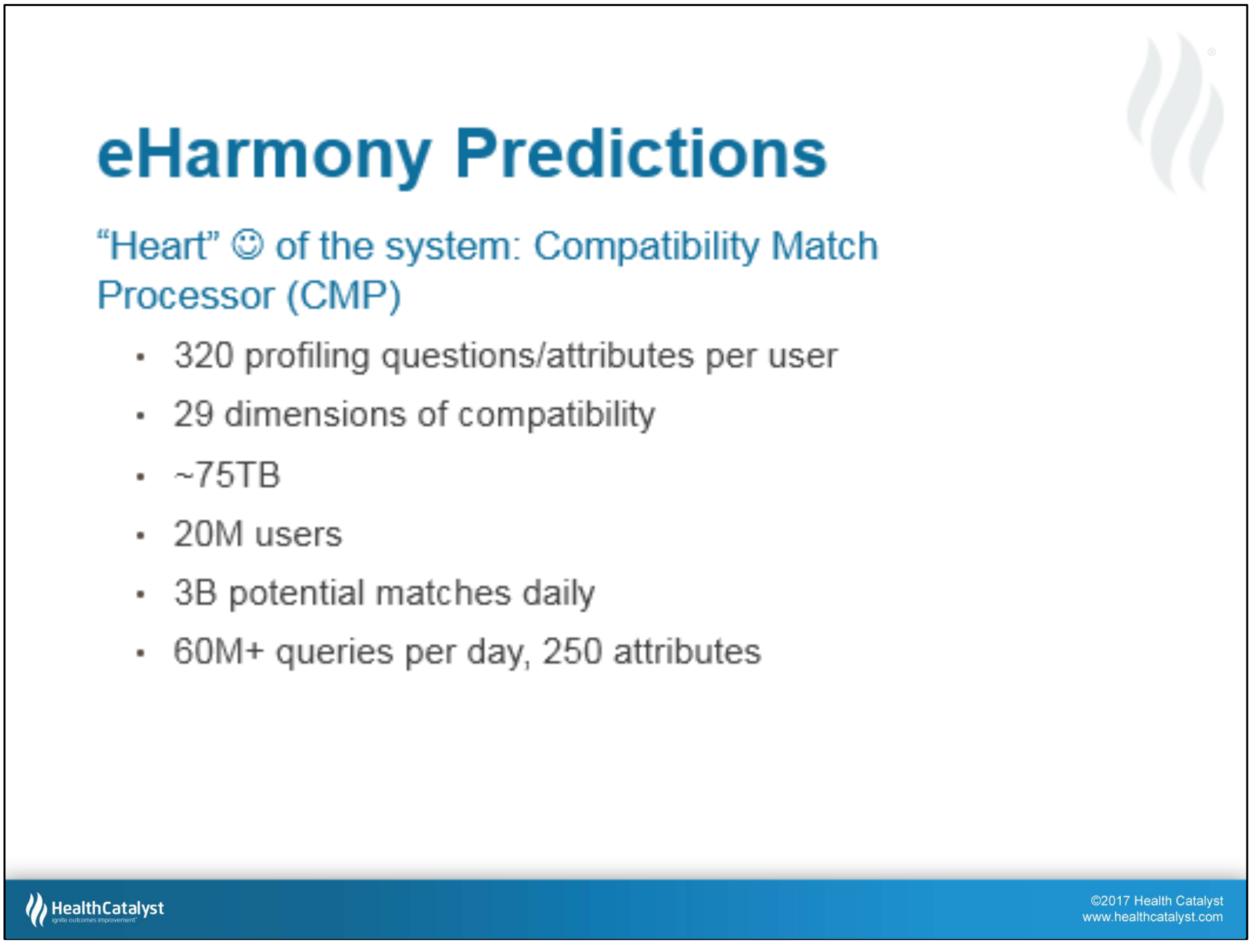
This is to give you some details, you know, insight into the kind of data that they look at. There are a number of scales. By the way, this is from a company in Canada who tends to be the leader in the world on this recidivism risk assessment analytics model. And they have 15 different scales that feed the predictive analytics algorithm. And within each one of these scales are a number of discrete data items that are collected that feed into that algorithm. And from a study at Hennepin County in Minnesota, they found that 42% of high-risk offenders under this model recidivate within 3 years.

So, kind of interesting, isn't it? I, a few years ago, didn't realize that this was already happening to us socially, and I don't, I haven't quite decided if I'm going to pass the judgment on or not but I find it interesting.



[38:43]

Earlier this year, there was a letter sent from the chair of the United States Sentencing Commission and expressing concern about this whole world of using predictive analytics in sentencing and corrections policy. I thought it was interesting that in that letter they referenced Moneyball. I don't know, those of you may or may not remember. We brought Billy Beane to our Analytics Summit in Salt Lake City and he gave a stunning presentation on the use of data for recruiting in the Oakland Athletics.



eHarmony Predictions [39:22]

But you can see now that there is suddenly some head-scratching going on, saying, you know what, we're not quite ready to accept the validity of these predictive analytics algorithms in the sentencing and parole management process. So, we'll see how this all turns out. And again, I think this is kind of foreshadowing what we're going to see in healthcare. As we start using predictive analytics more, some of these cultural issues are going to show up, including things like rationing.

So those of you who have been in the singles world will probably know that eHarmony has a predictive analytics engine and the "heart" of the system is their Compatibility Match Processor. There are some interesting metrics about what that looks like with 320 profiling questions/attributes per user, 29 dimensions of compatibility, quite a few queries and users per day.

29 Dimensions of Compatibility

Dimension	scale α	Male Score r	Female Score r	Combined Score r	Congruence Score r
1. Character	.847	0.494**	0.314**	0.497**	-0.282**
2. Self Concept	.774	0.449**	0.438**	0.532**	-0.142**
3. Emotional Status and Stability	.929	0.575**	0.625**	0.678**	-0.145**
4. Anger Management	.750	0.428**	0.418**	0.498**	-0.055(n.s.)
5. Obstreperousness	.845	-0.397**	-0.352**	-0.475**	-0.050(n.s.)
6. Family Goals	.770	0.213**	0.115**	0.184**	-0.158**
7. Family Background	.929	0.254**	0.218**	0.314**	-0.056(n.s.)
8. Intellect	.827	0.275**	0.169**	0.266**	-0.173**
9A. Energy: Activity	.887	0.144**	0.173**	0.184**	-0.072*
9B. Energy: Spontaneity	.822	0.156**	0.124**	0.183**	-0.083**
10. Spirituality	.889	0.288**	0.252**	0.288**	-0.176**
11. Education	.814	0.174**	0.154**	0.187**	-0.069(n.s.)
12. Appearance	.860	0.103**	0.198**	0.182**	-0.122**
13. Humor	.846	0.277**	0.218**	0.308**	-0.077*
14. Mood Management	.782	0.438**	0.363**	0.504**	-0.108**
15. Traditionalism	.816	0.373**	0.325**	0.375**	-0.198**
16. Ambition	.800	0.086*	0.116**	0.126**	-0.031(n.s.)
17A. Sexuality: Romantic Affect	.870	0.451**	0.343**	0.484**	-0.260**
17B. Sexuality: Physical Effect	.802	0.105**	0.163**	0.157**	-0.124**
18. Artistic Passion	.858	0.150**	0.108**	0.156**	-0.010(n.s.)
19. Values Expression	.829	0.212**	0.180**	0.226**	-0.016(n.s.)
20. Industry	.711	0.225**	0.176**	0.252**	-0.082*
21. Curiosity	.743	0.285**	0.158**	0.277**	-0.082*
22. Vital	.836	0.414**	0.368**	0.453**	-0.205**
23. Closeness	.828	0.497**	0.394**	0.513**	-0.259**
24. Communication	.881	0.502**	0.322**	0.499**	-0.333**
25. Conflict Resolution	.845	0.420**	0.367**	0.486**	-0.192**
26. Sociability	.856	0.142**	0.156**	0.185**	-0.106**
27. Adaptability	.702	0.443**	0.331**	0.473**	-0.182**
28. Kindness	.873	0.514**	0.322**	0.499**	-0.378**
29. Dominance	.757	-0.088**	-0.120**	-0.142**	-0.041(n.s.)

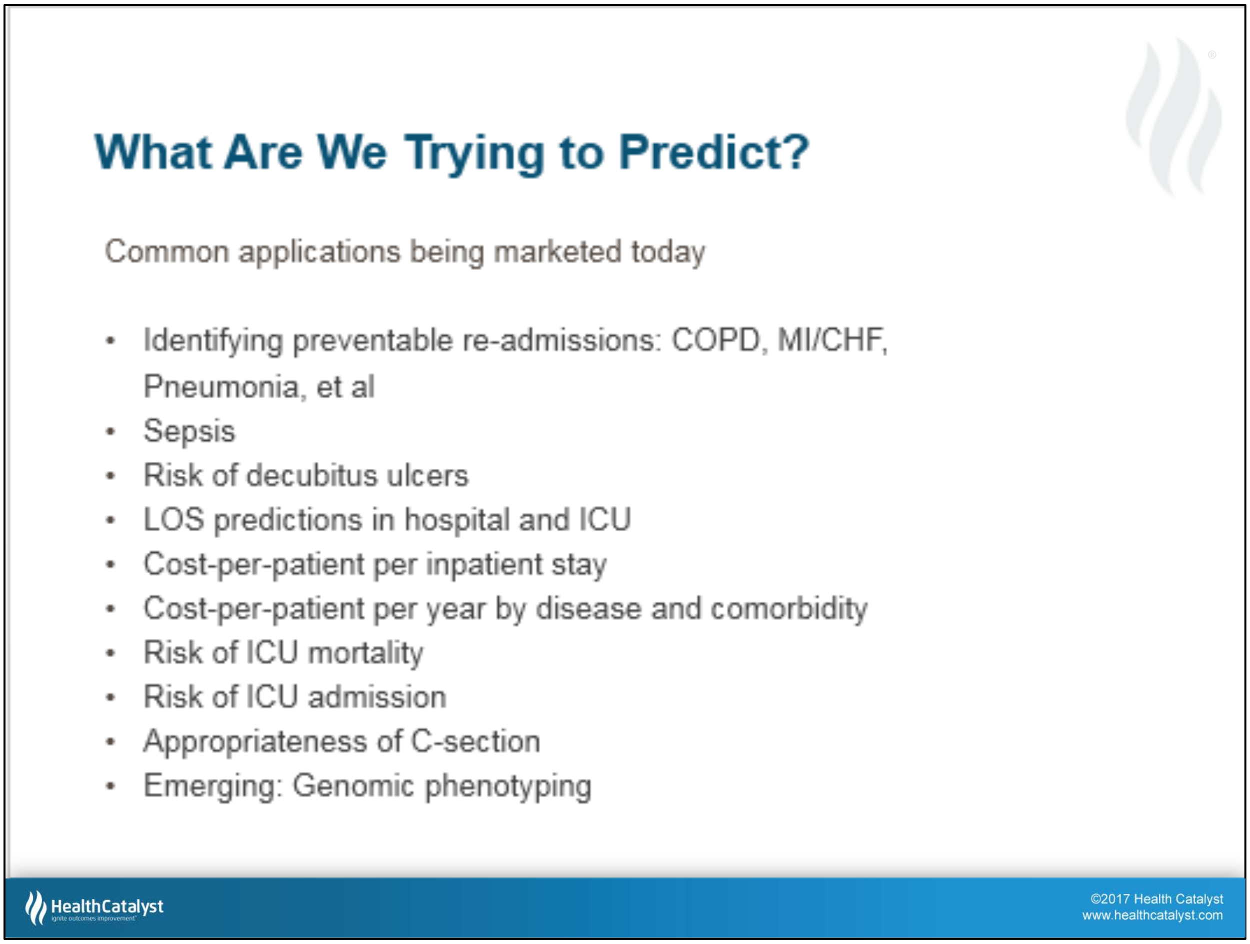
HealthCatalyst

29 Dimensions of Compatibility [40:24]

And there are the 29 dimensions of compatibility with all the outputs and correlation scores and all that. I won't go into this. Just to know that this thing is going on all around us, and you know, I'd be curious to know what did eHarmony consider a good outcome and how do they know if these algorithms are working. I see them advertising on TV the marriage rates. So that seems to be the outcome that they're tracking but to me I think there's probably a closed loop problem here. I don't think they really know what their outcomes from these predictive algorithms are. And if there's anybody out there that knows more about this as I do, I'd love to hear about it.

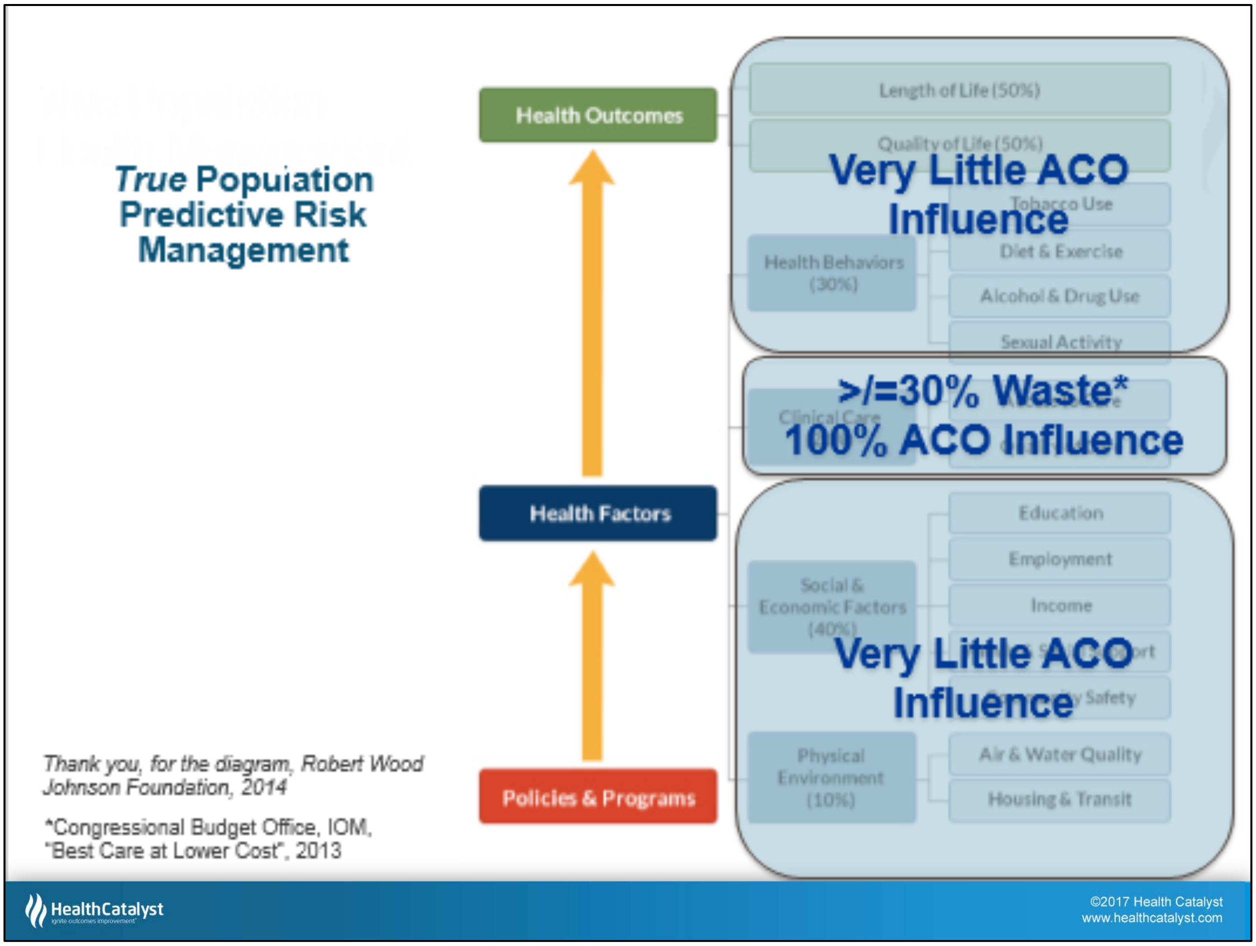
Predictive Analytics Inside Healthcare [41:05]

So let’s talk about predictive analytics inside healthcare for just a second.



What Are We Trying to Predict? [41:12]

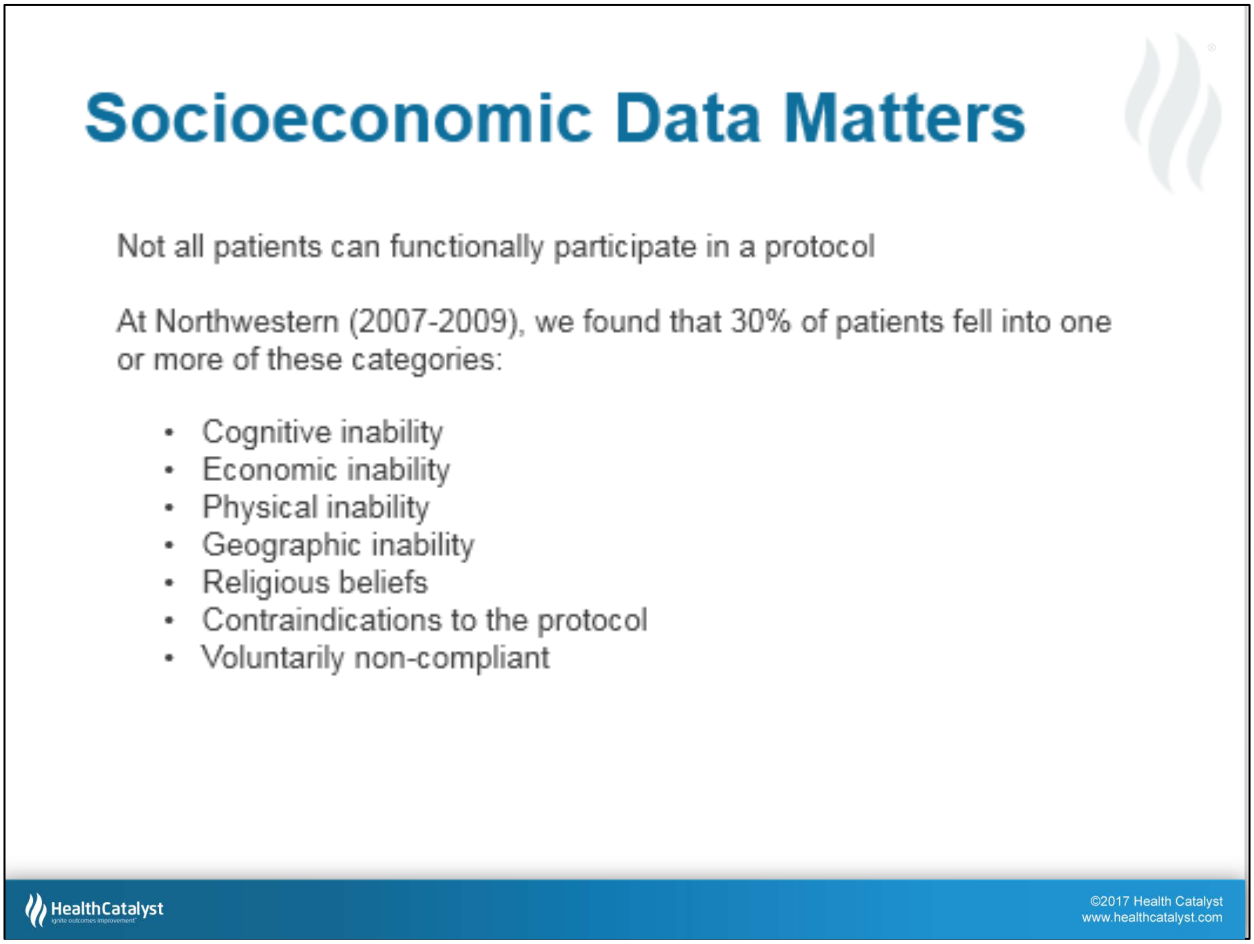
What are we trying to predict? There's all sorts of applications being marketed today. Readmissions is probably the most popular sepsis. And by the way, you can kind of divide the world of predictive analytics into acute and chronic condition management. So, a chronic condition management in a readmission state sort of crosses the boundaries a little bit. But generally those are the two worlds. It's either acute care monitoring and risk prediction and management or it's chronic condition and population health monitoring. And these are some of the examples. Probably not a big surprise to everyone. And I would say, you know, we've seen a lot of success on an inpatient side and I am again very grateful to be associated with Intermountain where some of the most forward-thinking uses of the health system, as well as applications that we would write around the periphery of the health system had a big impact on inpatient risk reduction and especially in ICUs and that area. So, lots going on, and again, this is the sort of thing that kind of tempers my cynicism. There's plenty of value to all of this.



True Population Predictive Risk Management [42:26]

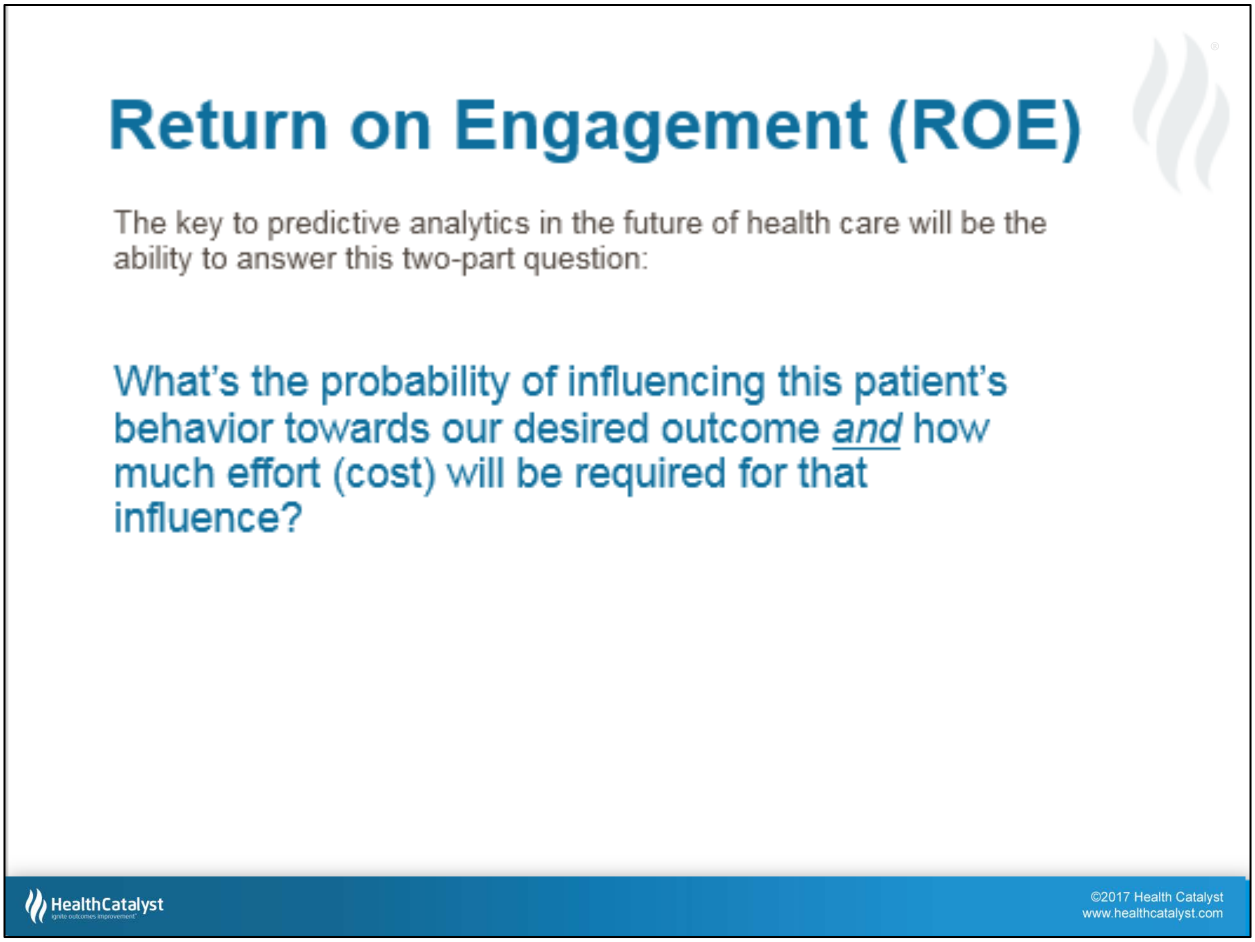
I would like to mention my slide isn't animated here but hopefully you can read through this. This is a slide that I love to mention from the Robert Wood Johnson Foundation. And what they described in this slide are the factors that affect healthcare in true population health management. If you can read through this, clinical care right there in the middle only accounts for about 20% of the cost and the risks associated with healthcare delivery. So what we traditionally think of in the provider space, we're only really affecting about 20% of the total picture. Health behaviors account for 30%, socio-economic factors account for 40%, and the physical environment the patients live in account for 10%. So all total, 80% of what really affects population health management, lies outside the boundaries of a traditional care delivery organization. And this is kind of what returns me to being a little more cynical because what we're finding, you know, in the ACO space especially, is that it's really hard to turn things around once those patients leave the four walls of your organization and they're out living their lives and they're susceptible for the influences of these other things, socio-economic factors, physical environment, drug and alcohol use, diet and exercise. Our ability to influence and predict what might happen based upon the variables in that space is pretty hard right now, in part because we don't have much data about what's going on in those areas.

So, again, my emphasis here is take advantage of what we know is significant waste and eliminating that waste to virtually every healthcare organization. There's all sorts of time and money wisely spent fixing that problem while we try to get our arms around what we're going to do in the future of population health management in these areas that lie outside our traditional areas of influence.



Socioeconomic Data Matters [44:38]

So we found at Northwestern that this socioeconomic data really matters. We found that about 30% of patients fell into one or more of these categories. Sort of challenging situations for patients to participate in their own care. They are cognitive inability, economic inability, physical inability, geographically challenge, you know, they're out in the middle of the farms and things there with little access to care. Sometimes there were religious beliefs that stood in the way of participating in protocols. But all of these things right now are for the most part not captured. We really don't know much about this. And so, we can't tailor our care to address these factors. But according to Robert Wood Johnson Foundation, they probably account for about 80% of the total healthcare risk in the country.



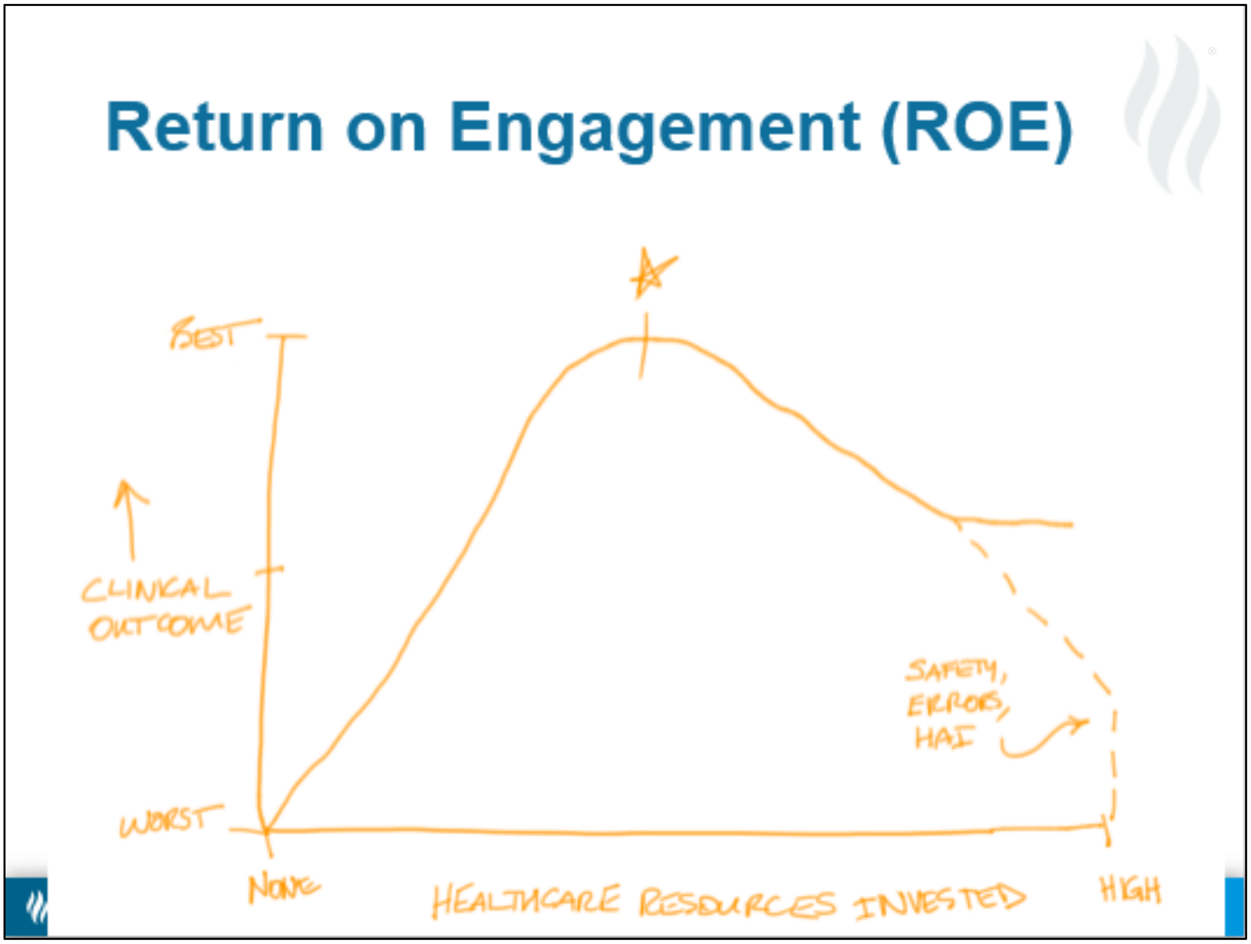
Return on Engagement (ROE) [45:39]

So I believe that this constant return on engagement is going to be very important in the future and it's going to be key to predictive analytics but it's also going to raise all sorts of cultural issues, like what we're seeing outside of healthcare. And the fundamental question is what's the probability of influencing this patient's behavior towards our desired outcome and how much effort/cost will be required for that influence?

So it's a different kind of prediction now. It's not about predicting an outcome, it's predicting the patient's ability to be influenced. Different kind of predictive analytics but critically important because so much of what we are trying to do is based upon our ability to influence those patients.

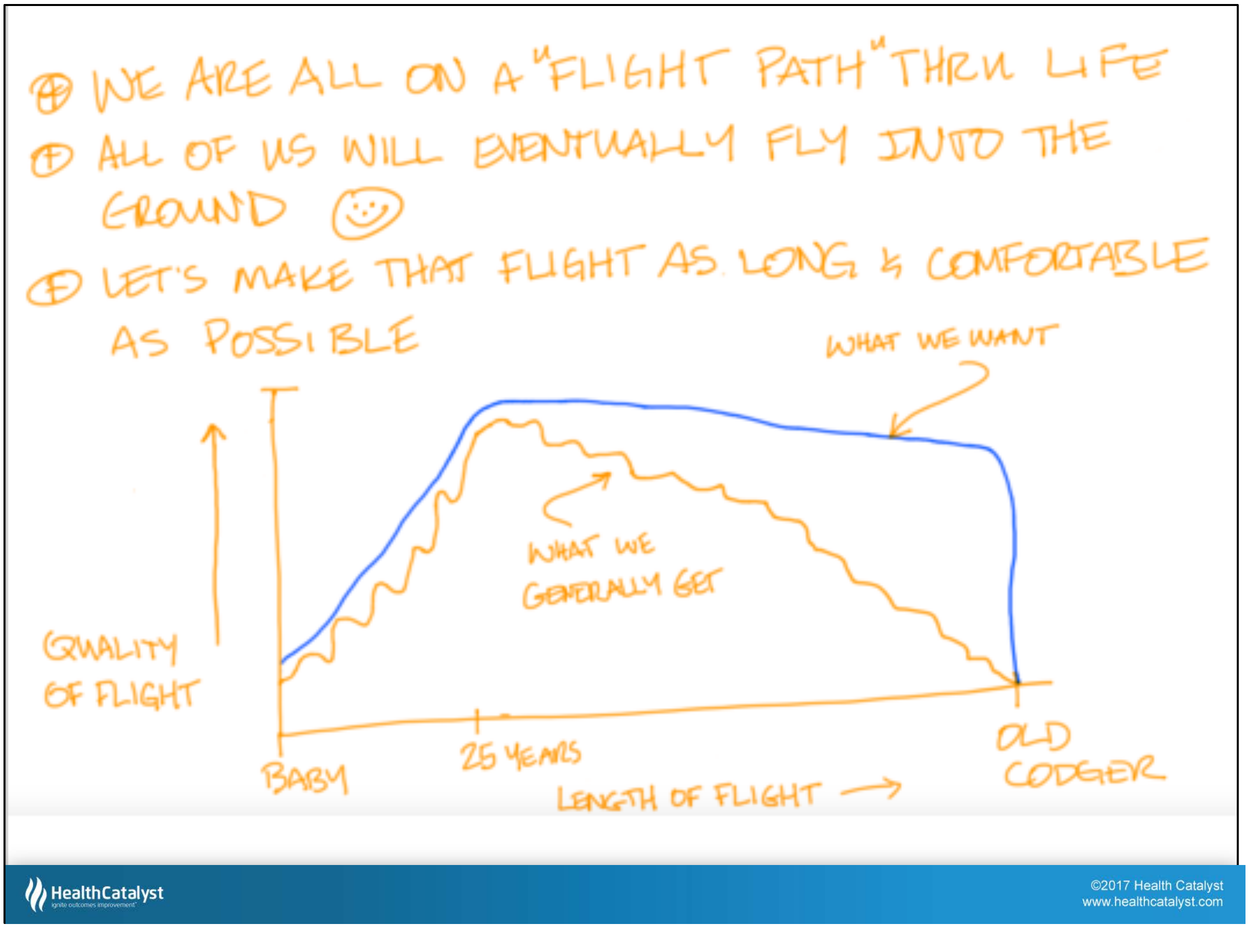
We found a real challenge, when we ran these algorithms against our obesity patients, we found that our high risk, you know, the algorithms produced the high-risk patients for sure and we could see those high-risk patients, but we also found that it was really difficult to achieve any return on engagement from those high-risk patients, that the best bang for the buck was for patients that were less risky from an algorithmic perspective. They were not so far down the obesity path that they couldn't be affected and turned around.

So you kind of have to ask yourself, well if I'm an obese patient, I don't want to be excluded by these concepts. I don't care if you think I'm helpless or lost cost or not. I want to know that I'm going to be cared for. But I'm not so sure that we're prepared to deal with those questions quite yet, and I think predictive analytics is going to reveal a lot of this. So we have to start thinking about now.



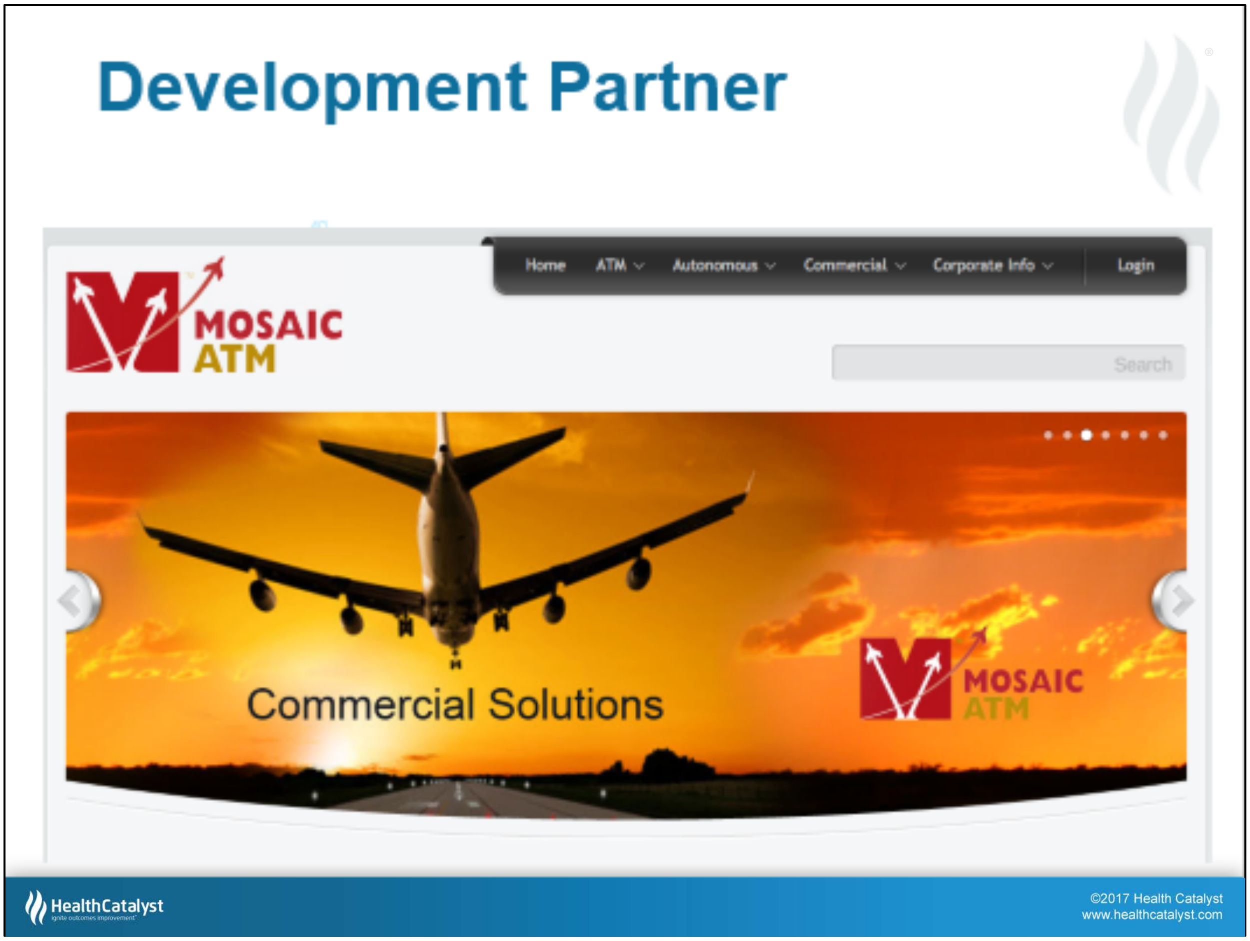
Return on Engagement (ROE) [47:31]

So, Return on Engagement is kind of this optimization of this curve, where your clinical outcome on this scale, you're trying to achieve the best clinical outcome, and balancing that with the healthcare resources invested, what I call the cost per unit of health. How much do you have to spend to boost this curve and push this curve up. Well, there's plenty of evidence that shows that you can invest more time and money and it actually reduces clinical outcome. So finding this optimal spot is what predictive analytics is going to be all about in the future.



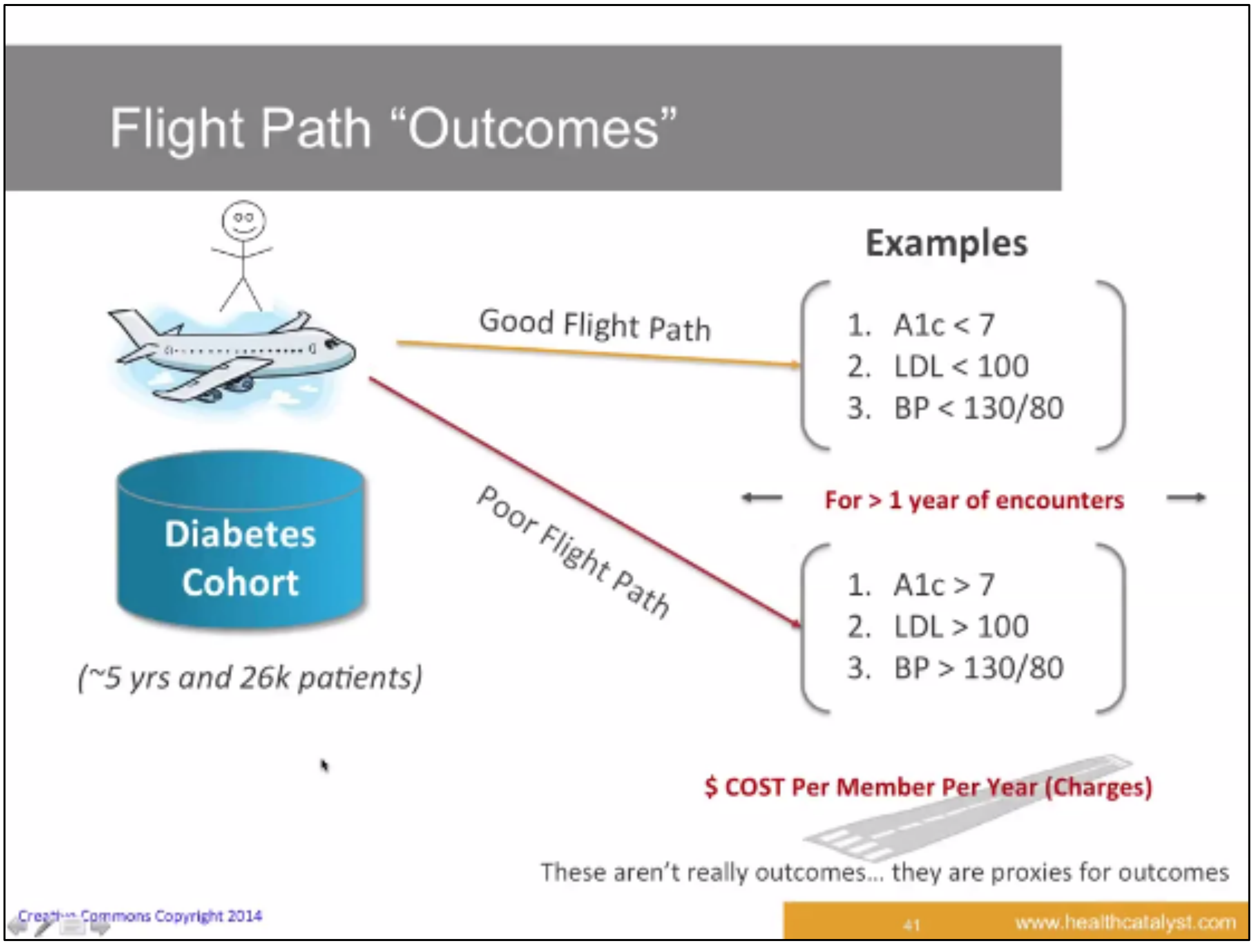
Flight Path [48:08]

We borrowed a concept from the Air Force about "flight paths". So I would suggest that we're all on a "flight path" through life and all of us are going to eventually fly into the ground. We know that. But let's make that flight as long and as comfortable as possible. So, if this is the quality of flight on this axis and this is the length of flight on this axis, what we're trying to do is get from being a baby to an old codger as smoothly and as long as we possibly can, stretch that out as far as we can. You could argue that our health kind of peaked to 25 years but it's this sort of (48:44) jagged up and down journey where we have little setbacks, where we come back and do various kinds of intervention in our body's ability to heal itself and respond. What we're trying to do though, you know, I think we'd all argue, we'd like to get up to that level of 25 and then hang on to that as long as we can, and then, you know, my preference would be to let it all crash and die in my sleep at night all at once.



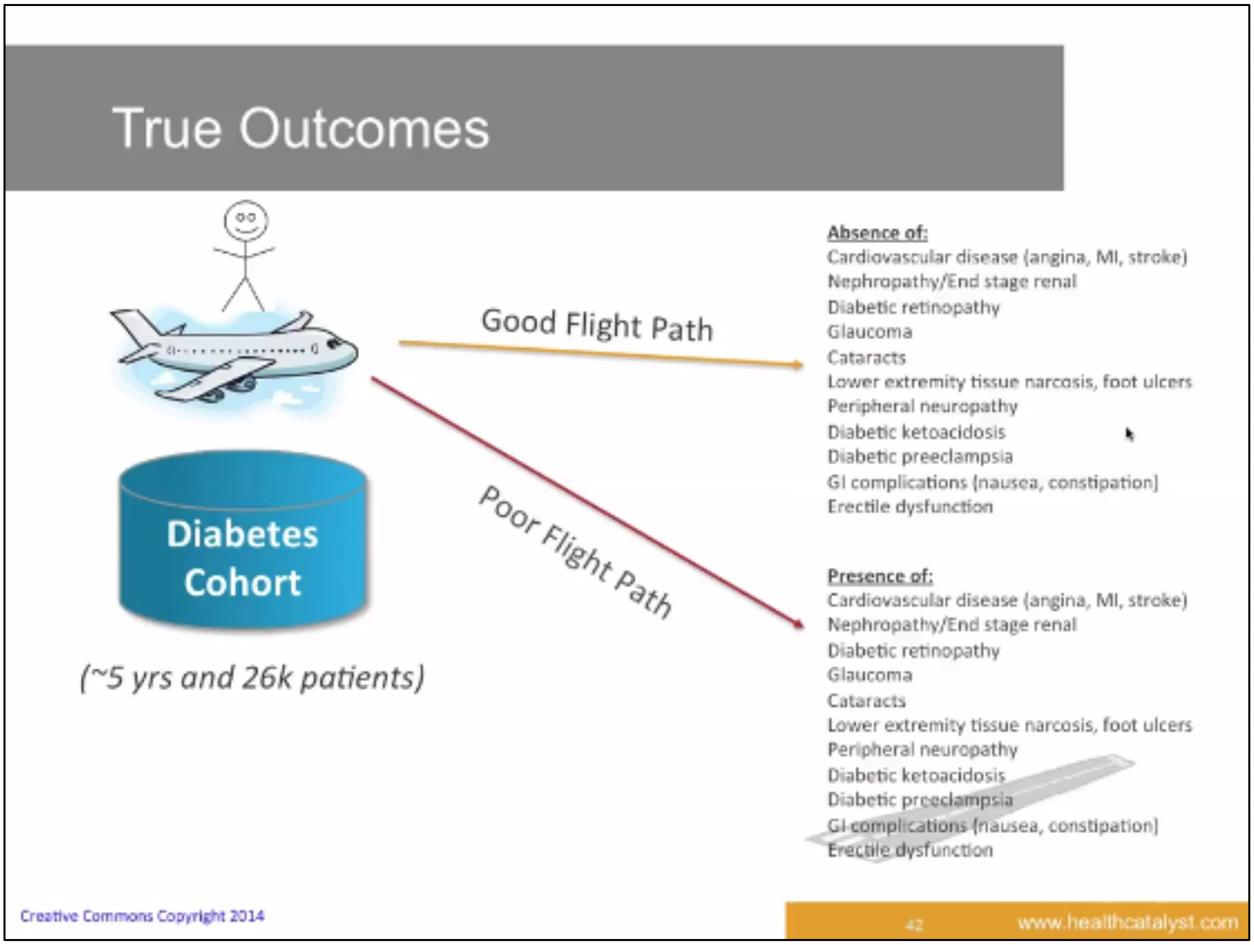
Development Partner [49:11]

So we borrowed this concept at a flight path and we partnered with a company called Mosaic, and they've been in the commercial airline, sort of predictive modeling world, for a long time. And again, this concept actually came from my work in the Air Force where we were trying to predict the turnaround time for aircrafts and what they call the order of battle management situation in tactical battlefield situations. So this notion of a patient "flight path" actually comes from that. We're partnering with these folks at Mosaic and we're working on the development of these flight path outcomes.



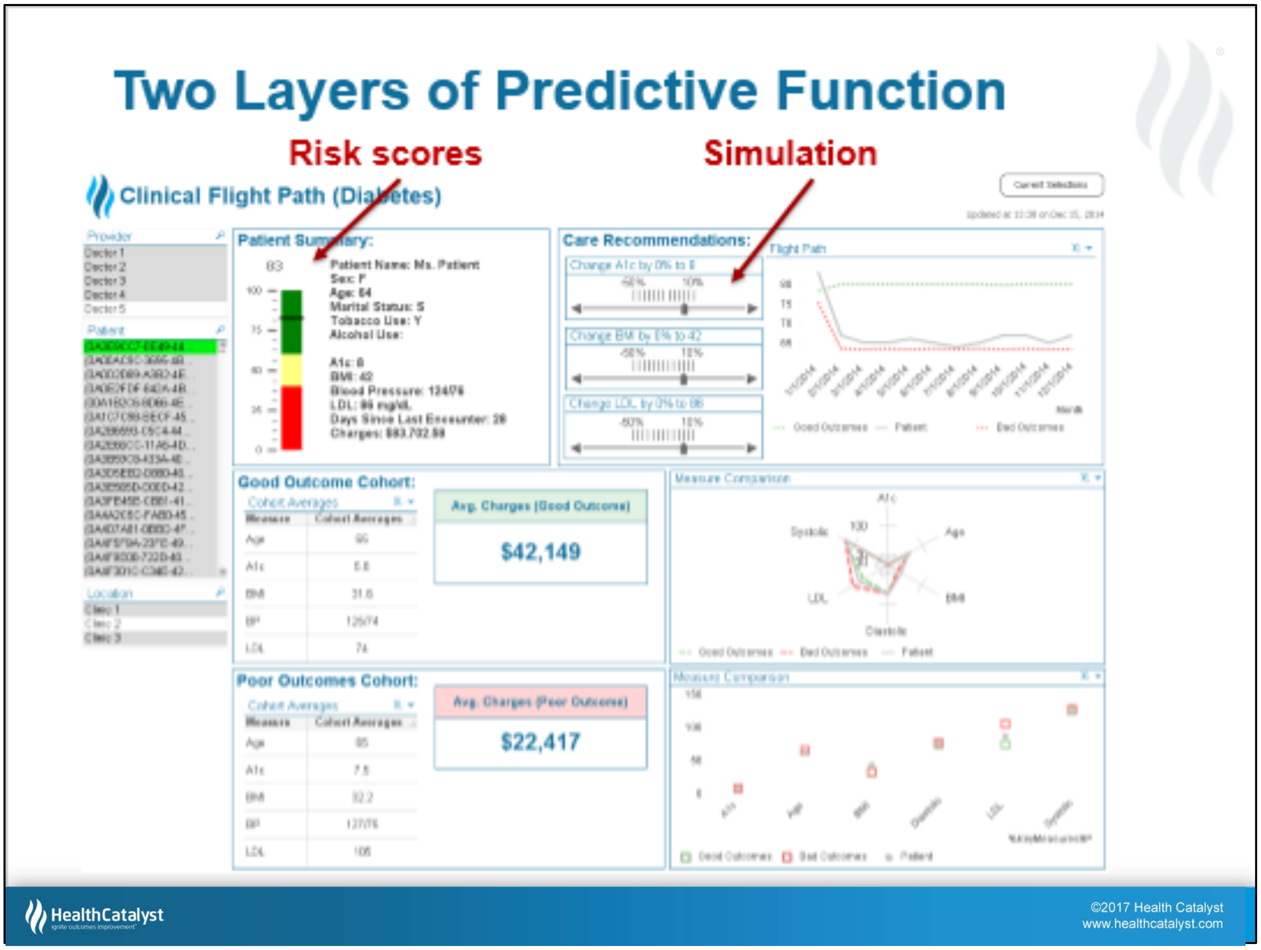
Flight Path "Outcomes" [49:46]

So, if we're diabetic patients, for example, a client site we're working right now, about 5 years of data, about 26,000 patients in the diabetes cohort. We're trying to understand those that have been on a good flight path and those that have been on a poor flight path. And these right now are kind of the example definitions of what a good and bad flight path is, as well as the cost per member per year. Now, I suggest that these aren't really outcomes. These are proxies for outcomes. We just assumed good outcomes by these observations but it's really not an outcome.



True Outcomes [50:22]

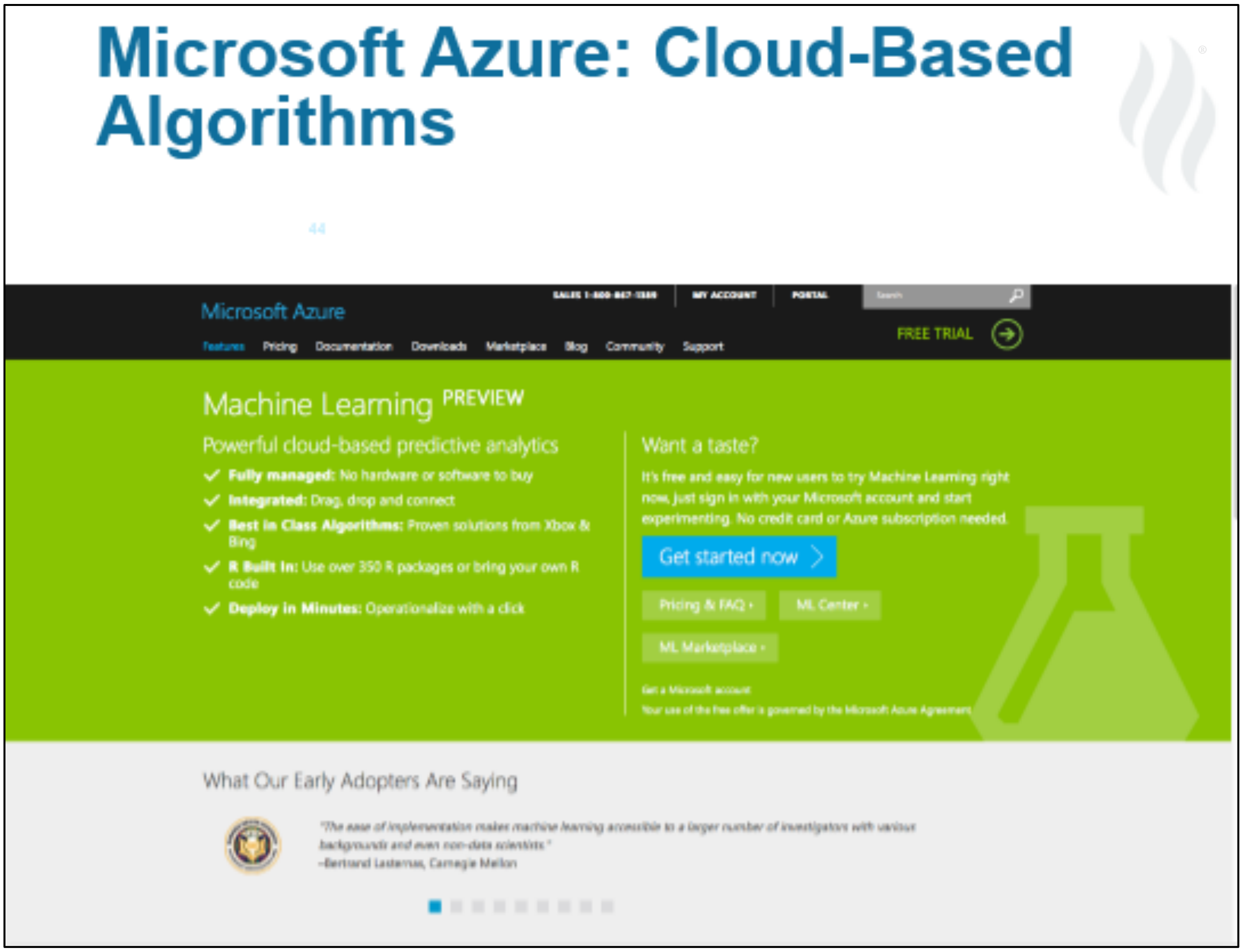
So, our second step is going to be more outcomes related, and that is what we're looking for is the absence of diabetic comorbidities and complications. My mom is a good example of a patient who had sky-high blood pressure for years and years and years, probably 30 or 40 years. But she really never had any downstream complications from it. She died at almost 90 years old in her sleep. So what we're looking for are not these proxies of sort of are you practicing evidence-based medicine controlling those patients, but what we really want to know are the absence of these complications from diabetes for the presence of, in the case of a poor flight path. And so, we're working on that actively right now at one of our client sites. I'm just very grateful to be associated with them.



Two Layers of Predictive Function [51:16]

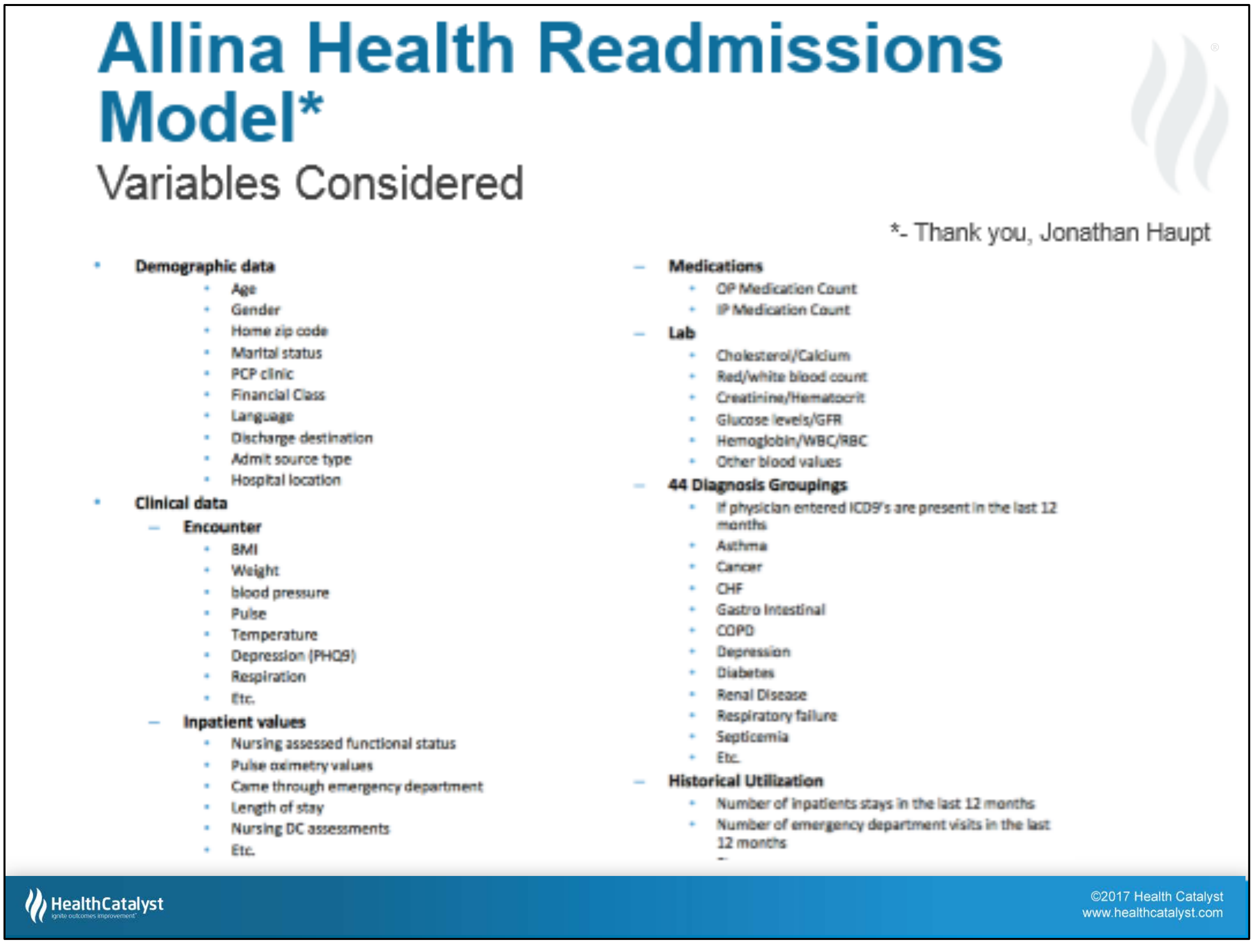
And this is a screenshot of that application that we're developing, and again I credit David Crockett and Brent Larsen and other folks doing all the work on this. And you'll see two kinds of predictive functions here embedded, risk scores for the patient, and you can select doctor, you can select patient over here, you can even select by location; and then some simulation, what if we could change this patient's A1c or BMI or LDL, what kind of effect might that have on their risk score and their flight path. This is their flight path area over here.

The other thing that we always try to do in all of our analytics is combine clinical data with cost data, trying to find that optimum point. And you might find it interesting, these are real numbers, still to be vetted but you know, it costs quite a bit to achieve a good outcome as opposed to a poor outcome. But are we spending too much? That's the question. We want to minimize that if we can. And what do we need to spend to achieve a better outcome for this patient?



Microsoft Azure: Cloud-Based Algorithms [52:23]

I want to draw attention as I did to Mosaic to what Microsoft is doing in the Cloud that makes machine learning very accessible. We're going to do some experimentation with them in this space. So now we're commoditizing access to these algorithms and things, which is awesome. So you don't have to be a data scientist or a data engineer to take advantage of it. And I would encourage everyone to go out to the Azure platform, sign up and start playing around a bit. It's really amazing what they've managed to make accessible to the general public. And as you may or may not know, Microsoft is kind of our core strategic partner.



Allina Health Readmissions Model*
Variables Considered [53:05]

Allina has a great readmissions model and I think Jonathan Haupt and Mike Doyle who were there and helped developed it. These are some of the variables that were considered in the development of their readmissions model. And by the way, these slides will all be available so you can look at those.

Allina Compared To Other Models

Multiple logistic regression

Model	LACE	CMS	Systematic Review*	Allina Health Readmission Predictive Model
Summary	4 variables: **L**=LOS, **A**=Acuity (was ED), **C**=comorbidity index, **E**=ED utilization	Claims based, many parameters	Varying depth and applicability	30 clinical and internal variables. Applies to all patients
C-stat	0.68	0.63 – 0.66	0.56 – 0.72	**0.73**

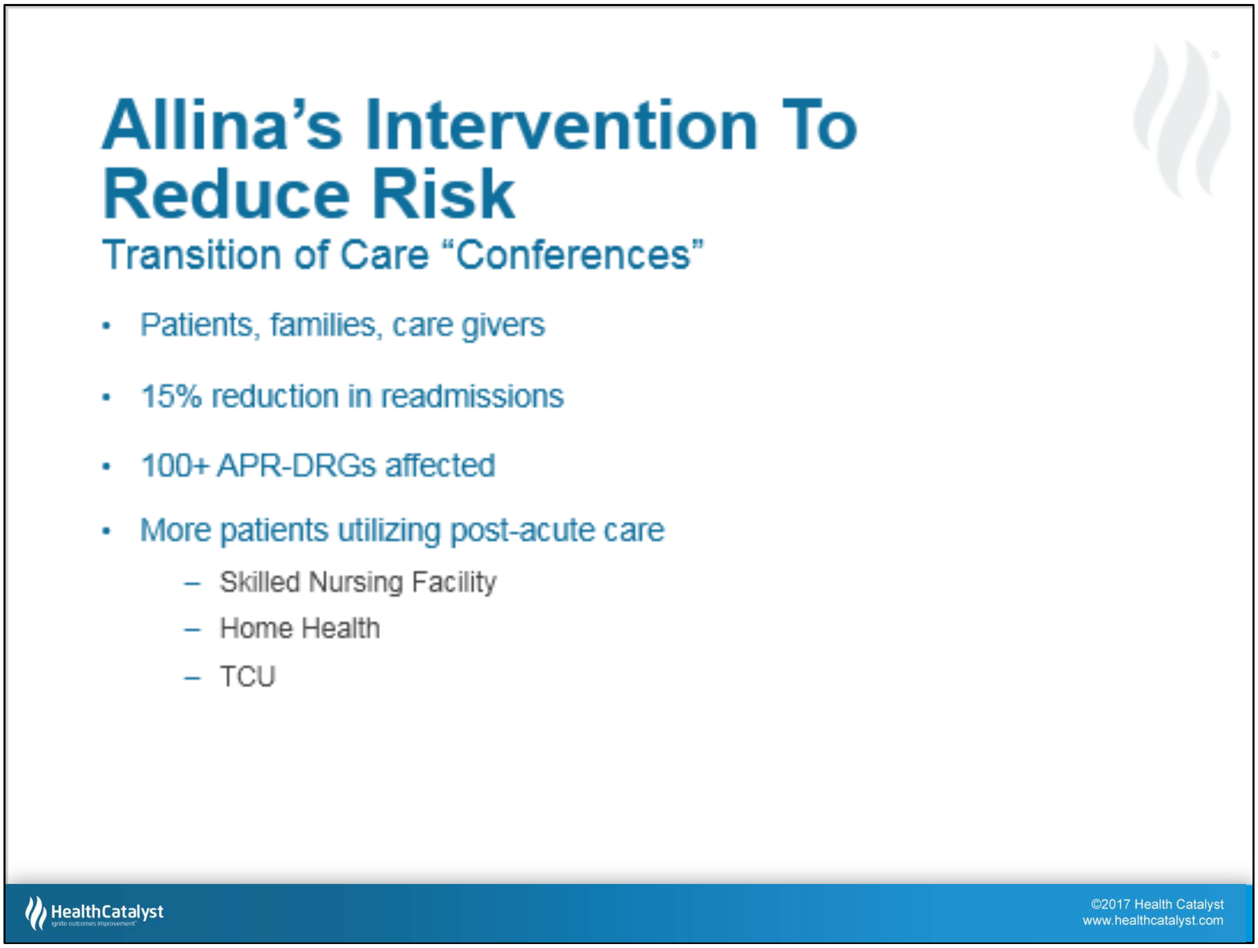
- *JAMA "Risk Prediction Models for Hospital Readmissions"
  - Oct 19th, 2011, Vol 306, No. 15, p 1688
  - 26 Unique models reviewed
    - 14 on claims data
    - 9 of those 14 had low discrimination ability (c-stat 0.55 – 0.65)
    - 7 with moderate discrimination available **during the stay** (c-stat 0.56 – 0.72)
    - 5 at hospital discharge (c-stat 0.68 – 0.83)
  - Vary widely between the groups (one will work great with Asthma but not AMI...)

5.2% of discharged patients in high risk category

HealthCatalyst

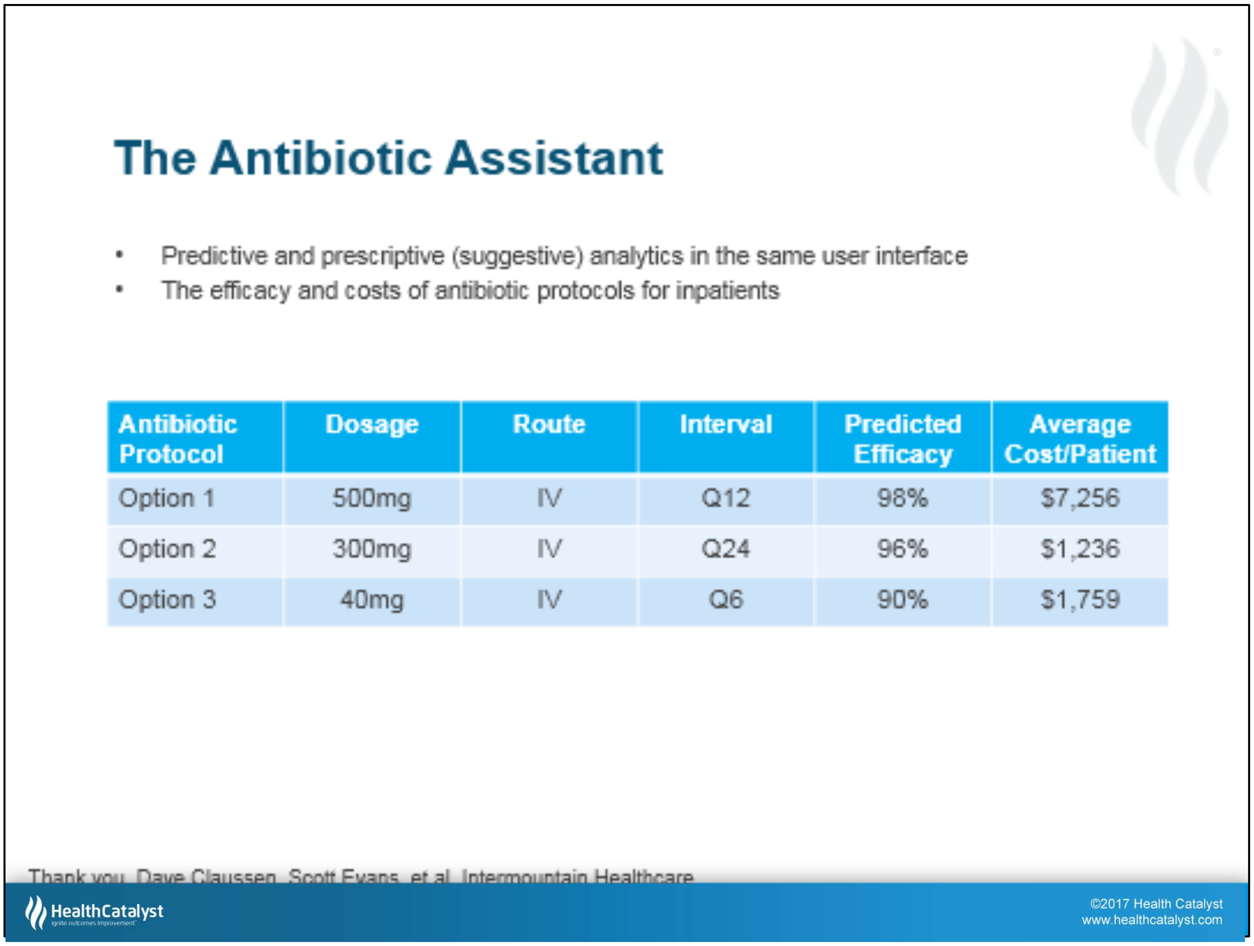
Allina Compared to Other Models
Multiple Logistic Regression [53:20]

In comparison to the other models that were out there at the time, which is now this is almost 2 years old, they perform much better with their predictive models that Jonathan developed. It resulted in 5.2% of discharged patients were in the high-risk category. And I might also mention that they used their multiple logistic regression algorithm for this and there's good reasons for that and they will talk about it at a later time.



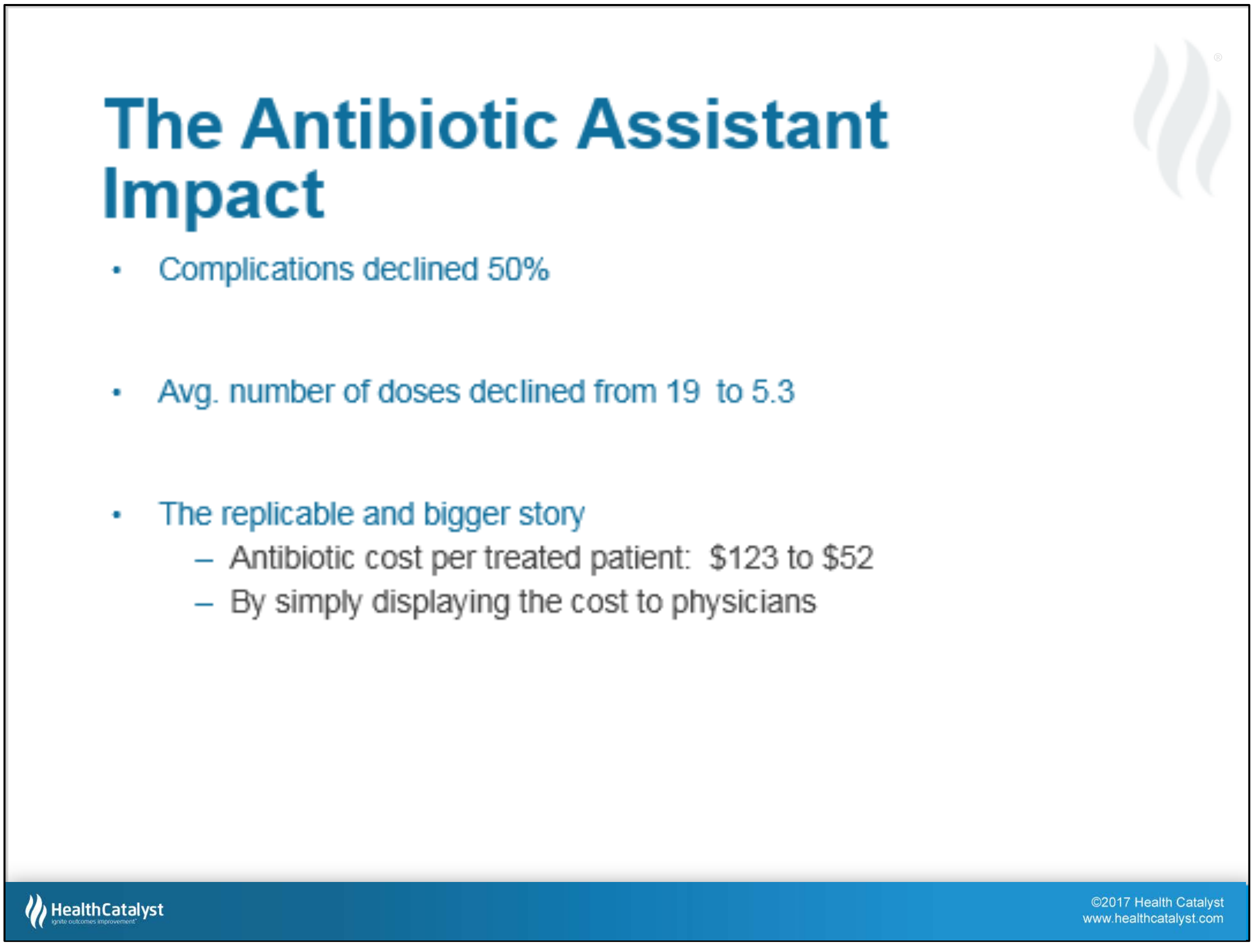
Allina's Intervention to Reduce Risk [53:55]

So what was Allina's intervention? Right? So they did the predictions. Now, what did they do to intervene? Well, they put together these transition of care "conferences" that included everyone involved, patients, families, and care givers. They found a 15% reduction in readmissions across more than 100 APR-DRGs and a greater utilization of post-acute less expensive care. So it's happening and it's working. And again, Allina is kind of our flagship Health Catalyst care site and partner.



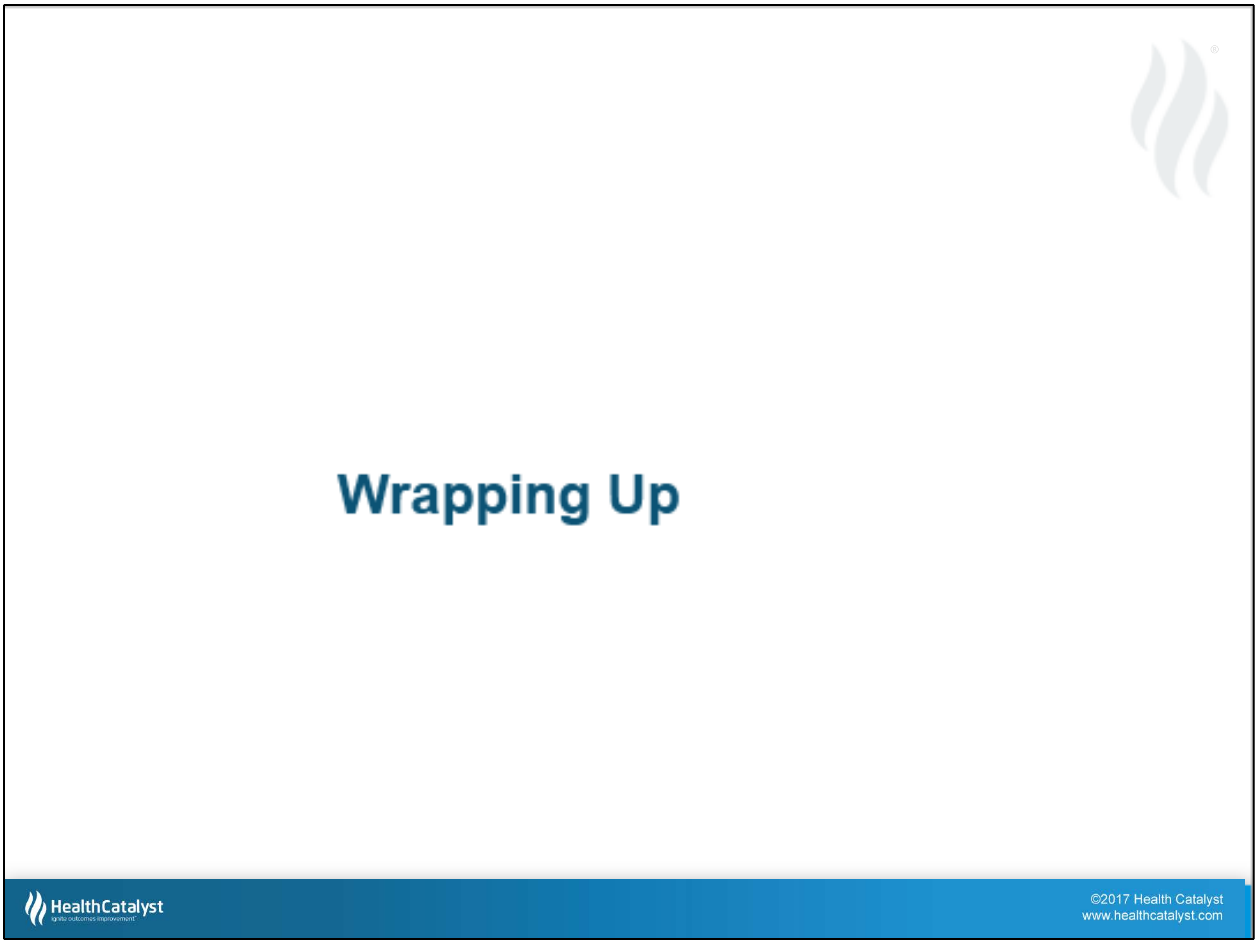
The Antibiotic Assistant [54:28]

I'll mention the Antibiotic Assistant real quickly at Intermountain. Again, one of those things I was happy to be associated with. Really grateful. Not only did Scott and Dave predict the efficacy, that's the predictive analytics there, but they were actually prescribing, right? So, they were predicting the risk of a patient but they're also prescribing what the algorithm thought was the best antibiotic protocol. And this example, even though it's specific to antibiotic management in an inpatient setting, it's really a framework for what I think decision support at the point of care should be. It's prescriptive and that it's offering protocols; it's predicted, in that it's suggesting what the efficacy of those prescriptive analytics might be; and it's also showing you the average cost per patient. So you're bundling the triple point or the triple aim right at the point of care.



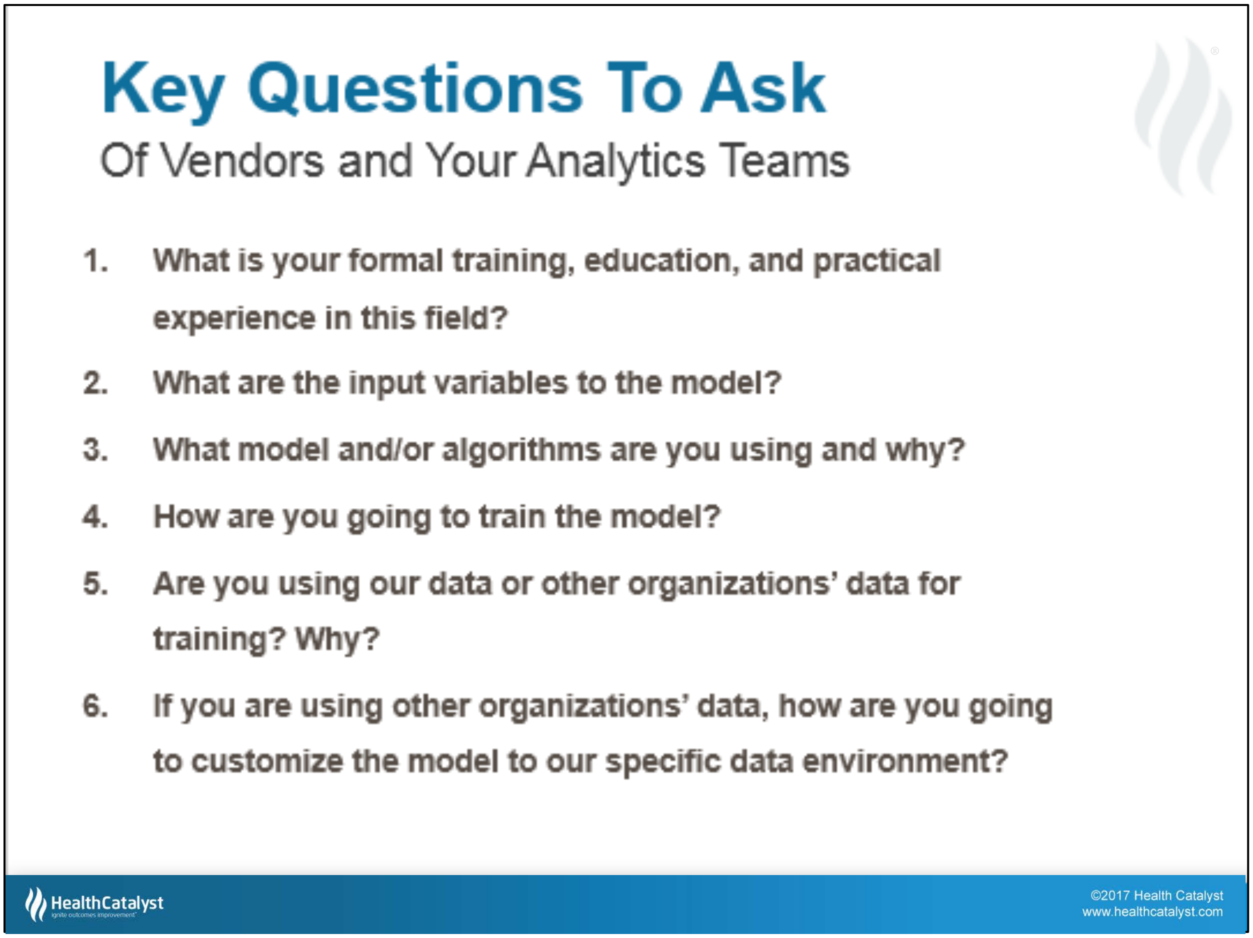
The Antibiotic Assistant Impact [55:25]

And there’s the impact. Complications down by 50%, number of doses slashed from 19 to 5.3, and cost per treated patient dropped from $123 to $52 simply by displaying the cost to physicians.



Wrapping Up [55:40]

So wrapping up here, friends.



Key Questions to Ask
of Vendors and your Analytics Teams [55:45]

Questions to ask of vendors and your analytics teams before you dive into this too rapidly.

First and foremost, what is your formal training, education, and practical experience in this field? Do you have the resume? Because this is not an area you want to get into and learn at the expense of the organization.

What are the input variables to the model that you're suggesting?

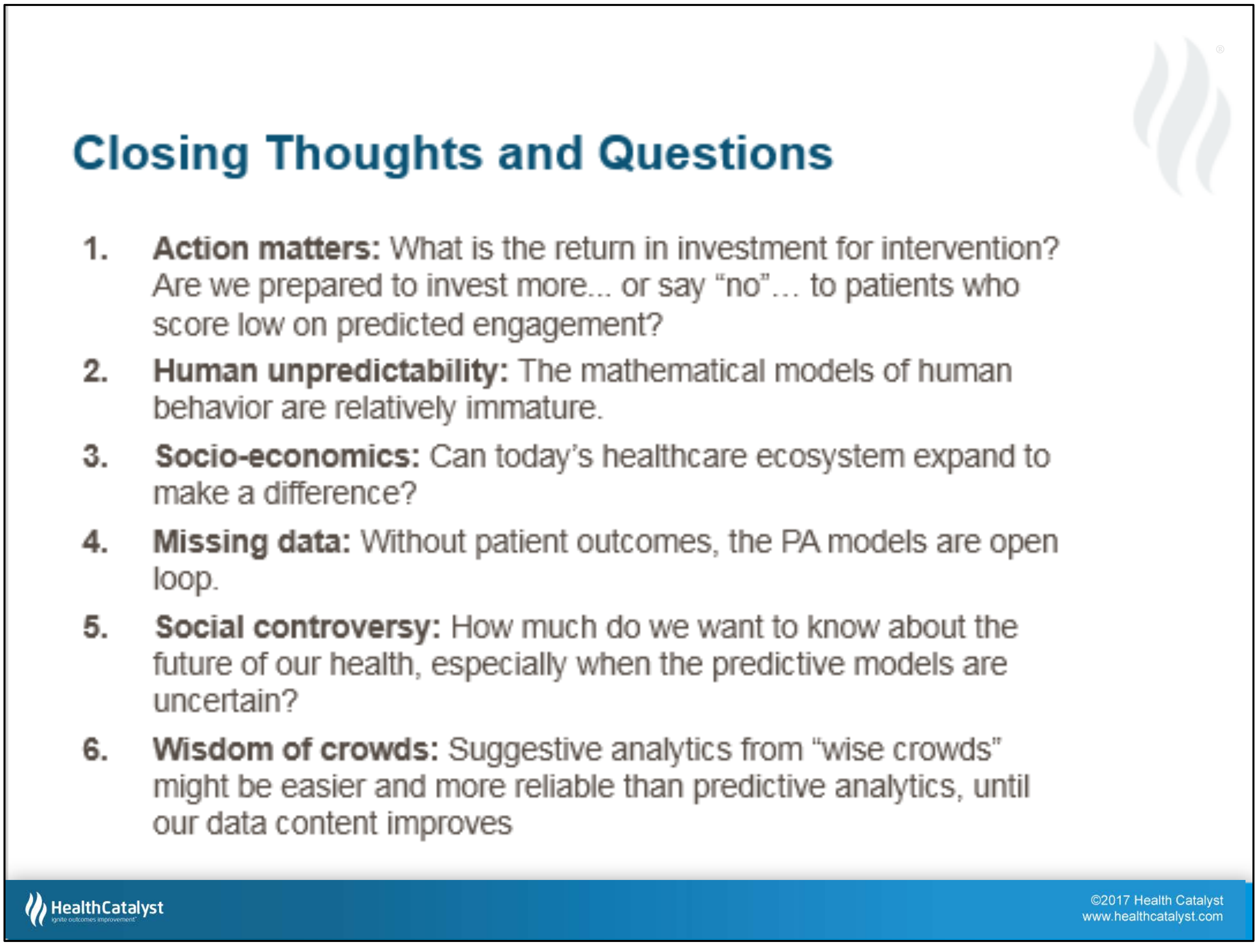
What model and/or algorithms are you using and why?

How are you going to train the model?

Are you using our data or other organizations' data for training? And why? And that's a big deal because especially in healthcare there's a lot of geographic specificity to data and it's best to train on your own data.

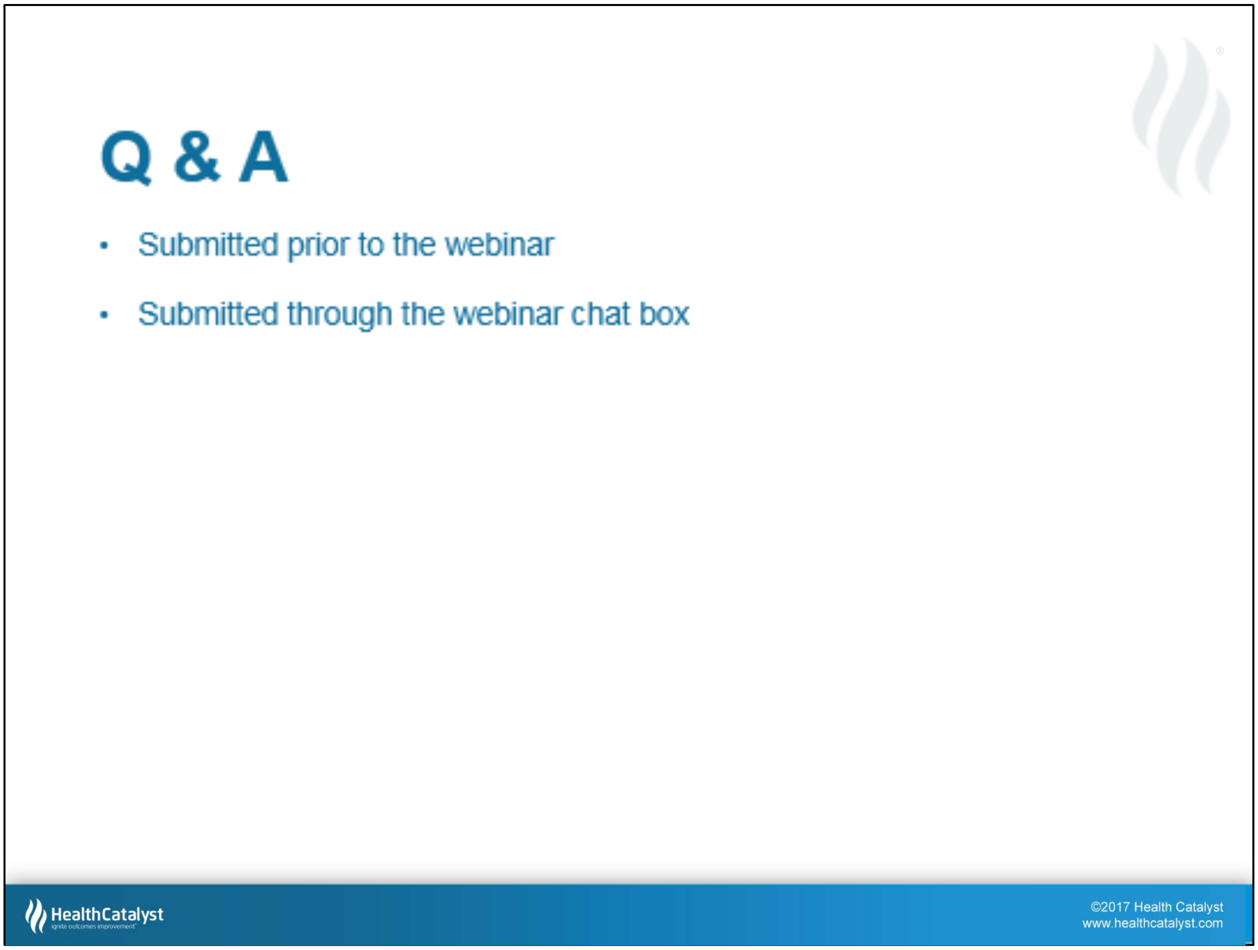
If you are using another organizations' data, Mr. Vendor, how are you going to customize the model to our specific data environment?

So those are some quick questions that will help shake out a lot of the problems and help everyone come up with a better solution in this field.



Closing Thoughts and Questions [56:52]

So, closing thoughts and questions. Action matters. Intervention is what it's all about. Humans are tough to predict. We're still quite a ways behind in terms of mathematical models. We're missing a lot of socio-economic data. We don't even have the ability to influence and intervene in those settings yet. We're missing data in healthcare. We've got to close that. There's going to be social controversy, so be prepared for that. And then we'll follow up later on this topic. The wisdom of crowds is very important, and what I call suggestive analytics is actually very easy in comparison to predictive that has very high value, and that will be a separate webinar.



Q&A [57:32]

So, there were some questions that came in prior to the webinar. And we're at the top of the hour. Now, I always stay past the hour. So I'm more than happy to do so, friends, and be glad to do that. Let me bring up the questions that were asked beforehand and I'm going to get out to my spreadsheet here somewhere. There it is.

QUESTIONS	ANSWERS
I am particularly interested in predictive analytics in neonatal care and what trends you may have seen in pediatrics overall. In this case, the parents are the third party receivers of the predictive information and we have some ethical responsibilities to consider.	I am going to plead a little bit of ignorance in this area. I don't know enough about predictive analytics in the pediatric space to speak real intelligently about it. That said, I don't see a whole lot going on outside of readmissions, but I will take that as a personal challenge on my part to become more familiar with this. And if you'll reach out to me, I'll help share what I learned from that. Actually, Health Catalyst has quite a number of Children's Hospitals in our portfolio. I just don't know what, if anything, they're doing with our predictive analytics to actually change care.
Are predictive analytics in your opinion useful in the	Well, I think they can be. Actually, the data is a little

pediatric context? To what extent.	different there, and again, one of the challenges there in the pediatric sense is the volume and type of data is lower than it is in the adult world. And so, there might be some unique challenges in that environment. I have to give that a little bit more thought.
How can predictive analytics be used at the point of care? This seems to be the clinicians' litmus test for usefulness of the data instead of the analytics which may be the driver for the hospital administration.	Well, you know, there are some organizations. Friends at Evanston North Shore up in Chicago, Jonathan Silverstein and a colleague there are doing some very cool things, Cleveland Clinic is doing some very cool things, Geisinger, where they're folding risk scores back to the point of care. And you saw the example that I gave with the Antibiotic Assistant. So, that's what I call closed loop analytics and we're going to write a paper on this topic shortly, but it's really important that we start bending these analytic algorithms back to the point of care to help physicians make better decisions faster.
I'm not sure if you were going to touch on this or not but I would be interested in how/when you envision PA tools being utilized by clinicians at the point of care.	Yes, I do. And I think certainly from my experience at Intermountain, we could see the value of those tools. I mentioned Antibiotic Assistant. We also used their glucose monitoring and adjustment program at the point of care in the ICUs. There was a predictive model. Of course, the kind of the famous one was the ARDS weaning protocols and that was all at the point of care. So, absolutely positively this is closed loop analytics at the point of care. Really important to our future.
I found a recent visual by the Advisory Board helpful in framing a few different levels of analytics.	That's probably Jim Adams and Aranow. My friends there at the Advisory Board, I hope.
I'm interested to hear what you think about the difference in terminology and if you think it's helpful to separate predictive and prescriptive in analytics discussion. I'd also like to hear about examples from healthcare about how organizations are using predictive analytics in a way that customers and patients appreciate versus feeling like they're being intruded upon.	Yes, I do. As cynical as I am about predictive analytics, I'm probably even more cynical about prescriptive analytics in healthcare because to be effective with prescriptive analytics, you have to have evidence-based guidelines about what you want to prescribe, and we have had a heck of a time in healthcare coming up with evidence-based guidelines that we can computerize and that everyone can agree on. So, I think we have a long long way to go to be effective with prescriptive analytics in healthcare. We've got enough challenges to get our hands around predictive analytics right now. So that's where I would focus most of our attention. A great question here about whether patients appreciate versus feeling like they're being intruded on. I don't think we know the answer to that yet. Our credit risk score is a great example of that. We've all

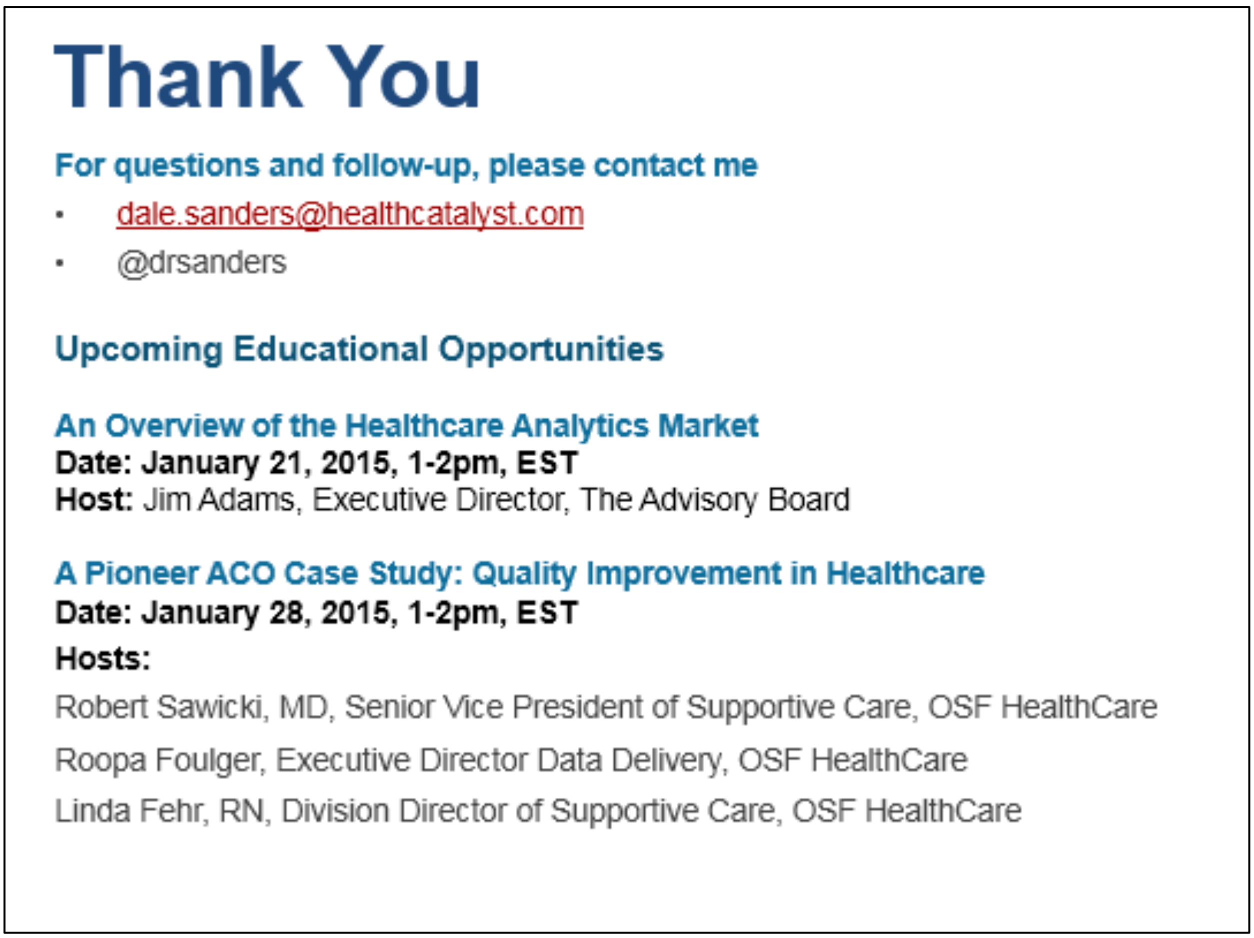
	come to accept that as a part of life in the United States but I would guarantee you that it has caused a lot of trouble for us on a number of occasions. So I think multiply the social complexity of that problem by orders and magnitude and I think that's what we're going to see in healthcare as predictive analytics goes forward.
It seems to be that predictive analytics is an effort to answer a question. I guess my question is how do we know what questions we need to ask to use PA effectively to achieve organizational outcomes and not just use it for prescriptive recommendations?	Well, you know, our methodology at Health Catalyst is all kind of focused around a really simple concept of the Pareto's Rule, and that is 80% of healthcare costs, risks, and problems is associated with about 20 to 25 care process families, and our godfather on this topic, Dr. Dave Burton, a colleague of Brent James at Intermountain, is the leading thinker on this topic. And so, what we would advocate is use predictive analytics to target those 20 to 25 care process families that constitute most organizations' highest risks, and it becomes pretty clear once you start thinking into it, how you could use predictive analytics, just like we are at the client site I mentioned. We're focusing on diabetic patient population right now and soon after that will be CAD and CHF patients. So, they're all pretty much the usual suspects. There's not a lot of variability across the high value opportunity areas in healthcare.
Are you going to talk about the regulatory environment and what is required for real-time clinical predictive analytics? I.e., FDA 510(k) and such? Also, how far away is predictive analytics? Who has implemented anything real and if not, when will we see real solutions?	I'm not aware of any 510(k) processors that regulate predictive analytics yet. So if you're aware of that, I'd love to know. What I see at the national level is less and less interest actually at the congressional level on regulating algorithms and kind of functional issues around an EMR in general. I think it will be interesting to see how ONC evolves around their focus on safety as it relates to EMRs. But I've been talking to the FDA and the 510(k) folks now for 20 years about this topic. I thought that they would be more interested in regulating this but so far they haven't. And by the way, I might mention, we have predictive analytics algorithms running on mothers during labor, and if you're crossing the border now, that's kind of the device-type function. So, I don't know. It would be interesting to see how this all turns out. I kind of think it's the patient that is ought to be more closely regulated but I'm also not a big fan of government heavy handedness.
I've heard some health systems talk about that they do not have big data yet, but are small or medium size data. Is there a right size of data for it to be beneficial for hospital reporting?	So I kind of touched on that. Generally speaking, the more data you have about your local population, the better. And again, I'll say that about your local population. So a lot of vendors are selling these

	comparative analytics platforms where they aggregate data from across the entire country. But I would suggest that that comparative data isn't all that valuable because there's so much variability geographically in patient demographics and there's so much variability in the delivery of care, when you start aggregating that data, the variability I would say is misleading in that you might believe that there's less variability than there is. So, the key thing is to aggregate as much data in your local population health management and geographic area.
There are many vendors that collect data in different databases. How can a hospital easily combine and normalize these different data sets?	Well that's kind of our job as Health Catalyst, is to provide that data aggregation, data warehouse platform, and that's what we think we do better than anyone in the country. It's quickly pulling those different data sets together and then laying applications over the top of them. And there's others in that space too but quite humbly we are the best at that by far.
Given the exploding types and sources of health-related data, what technologies and processes are available to extract meaningful predictive patterns and models – or are we doomed to live with the limitations imposed by the shortage of data scientists in the healthcare field?	Well, that's an interesting question. I do think that we have an enormous shortage of data literate, data capable people in healthcare and we need to ramp that up. We need to recruit from other industries and train them into the healthcare space. There are technologies like them, that Microsoft Azure Machine Learning Platform. It's starting to commoditize what you used to have to program in R on your own. But now Microsoft is making that available in the Cloud. So there is some disintermediation of that dependence upon these data scientist skills. It will be interesting to see how fast that moves and we're really anxious to work with Microsoft in that regard. And IBM Watson now has a Cloud-based tool as well for this kind of thing. So, we're starting to see the commoditization in the utilitarian use of these very sophisticated algorithms and models.
Given the anticipated growth in volume and importance of predictive analytics, what approaches will be needed for model governance, compliance and auditing of these models?	Well, that's great. As the data governance progresses, these predictive models will have to be governed in some fashion. So, there will become a data governance subcommittee that's focused just on, I believe, predictive and prescriptive models and algorithms and making sure that those things are governed appropriately so that they don't expose the organization to great clinical risks or financial risks. So most definitely the evolution of data governance will

	be towards algorithmic governance.
Do you anticipate that healthcare providers (like insurers?) will need to compete on the quality/reliability of their predictive models?	Most definitely. And we're already starting to see that a little bit in the healthcare organizations that are already taking on bundled payments and things like that. This is most definitely the future as healthcare providers start to look more like insurance companies and taking on that financial risk. Their ability to use predictive analytics and analytics in general is going to be a big competitive differentiator.

[Dale Sanders]

You know, I realized I didn't show a last slide here, friends, about upcoming webinars. I'm not sure how many folks we still have on. We've got 150 folks still on the…Let me bounce out really quickly before I take in more questions and just show another slide. Let me show this to you really quickly.



Upcoming Educational Opportunities [69:48]

So I needed this to tune me up. Here are some upcoming educational opportunities, *An Overview of the Analytics Market*. Jim Adams of The Advisory Board is going to provide that for us. And we're going to go into *Pioneer ACO Case Study* sponsored by our friends at OSF who are also a client of Health Catalyst. So I encourage all of you, if you are interested, please participate in those webinars. We surely enjoy your participation.

Questions and Answers

Okay. So I'll answer questions for about another 10 minutes here.

QUESTIONS	ANSWERS
For proprietary data types, however, there is much public data that can be mined such as clinical trial populations, what is your approach towards such data?	That is a great question. We don't have a great strategy around publicly minable data right now. We use it in our product development actually. So a lot of the states, in particular California and CMS at the federal level make these data sets available and we use those to develop our products quite often. We are not using those that I can think of in a client setting yet other than death registries. That's pretty commonly used. But you bring up a good point. We don't have a great strategy around how we might leverage this publicly available data to help some of our clients more locally. So, good question. Thank you.
In what area of healthcare you see predictive analytics becoming mainstream first?	Well I would say it's mainstream in a certain few areas on an inpatient setting already. Sepsis scoring, APACHE scoring in the ICUs, false risk scoring to some degree, readmissions is becoming pretty mainstream. You know, gosh, I don't know. I'm going to think a little bit more about that, but I'd like to say that chronic condition management, diabetes and congestive heart failure, even cancer. But the challenge there is back to what I mentioned earlier and that is we just don't have data outside of the care delivery system about the patient's lives in those socio-economic areas. So, yeah, I think the mainstream areas are those that I just mentioned and I'm going to give that a little bit more thought about where I think it may hit first outside of those areas. Thanks for being thought-provoking.
Will wearable devices play a meaningful role?	I think so, yeah. Yeah, we're starting to see that a little bit. I don't know that the kind of data that's being collected right now is going to make a big deal. I mean the kind of data that's more important is kind of at the blood level. It's laboratory-based data. So, figuring out a way to do the Star Trek diagnostic test by pressing a device up against your skin without actually puncturing the skin is kind of the – I think

	that's when we'll start to see a big change. And of course, you know, there are people that are wearing heart monitors and things at home right now, and they've been doing that for years, you know, that will dial in and send an alert to our care provider team and that kind of thing. So yeah, that's a short answer there.
Those promoting behavioral modeling population engagement plan that's greater than 30% of healthcare under control of the individual. So, if we just get them to change their behaviors, we can improve health. Why couldn't an ACO tend to further that?	Well that's kind of the fundamental question. I think we all have to have more economic skin in the game before we're going to be more interested in controlling our healthcare costs. I injured my wrist mountain biking not long ago and I've been battling the repair of that wrist for a while. I went to a hand therapist locally and found out, you know, she made me this custom brace that was $200 and billed the insurance for it and (74:26) we can do about it. When that brace wore out, I went to Walgreens and bought one for $17. And so, until we start becoming more engaged at a first order economic level in our healthcare, I'm just not so sure how motivated people are going to be to participate. And then again, what we found at Northwestern is that unfortunately there's just a lot of folks who can't participate. They have cognitive issues, language barriers, their social environment is not supportive. So there will always be a large number of folks who just can't participate actively in their own care and we have to decide as a society to what degree are we going to be compassionate and help those folks.
Flight path is similar to rectangularization of the death curve.	Yeah. Thank you. I appreciate that, friend, from BCU. I should have mentioned that. Thank you.
Are you using existing diabetic models as part of your work on locating diabetic cohorts?	Not so much. We're trying to do this completely innovative without being influenced or biased. So we may have to pause and go back to the existing models but we're not right now. We're developing something that we think is completely new and again borrowing from the brains and skills of the folks at Mosaic who have been working on this kind of things in the airline industry for a long time.
What regulatory legislator or court ruling impediments do you see now or expect in the future that would slow or stop mere experimentation in or actual deployment of either predictive or prescriptive analytics?	Well, this will sound kind of interesting. I think one of the biggest impediments is sort of ironic and that has been the meaningful use incentives that de-incentivized innovation in the healthcare IT market space. Suddenly, when we had $25 billion in federal money dangled in front of the market, we eliminated any incentive for free market creativity to step in.

	So I think one of the biggest regulatory and legislative impediments to innovation has actually been the high-tech investments that we've made. We could have managed that a lot differently. Other than that, I don't see anything. I would be interested of other folks who have seen it. I don't see anything happening at the congressional or the ONC level that would stop this. Now, HIPAA, you know frankly we over-apply HIPAA, I think, quite often in healthcare and you could call that an impediment. And if you know me, I'm very paranoid and respectful of patient identities and patient matching, but the way we've implemented and the way we interpret HIPAA is definitely standing in the way of experimentation and innovation. What will be interesting to see, I think, is whether Institutional Review Boards, the IRBs, to see what kind of role they play in this. My experience with IRBs in this area has not been very positive. They tend to be more harmful than helpful. And so, if you've got a predictive or a prescriptive analytics program, I personally would be less inclined to engage with the IRB. Keep it a little quiet. Obviously you don't want to be doing patient harm, but a lot of this can be experimented on in data and there is no harm to patients. So, I don't really see anything happening right now that's going to stand in the way of innovation.
	Yes, we will be posting this and making the slides available. Thank you.
Do you think consumer purchasing data groceries and suggested alternatives consumer choice will be something we will see in the future?	Well it's actually already happening very quietly at a handful of places. So we can go out and buy healthcare purchasing data from Acxiom and Experian right now. And in fact, David Crockett on our team has been talking to Experian about doing that. So it's going to happen. I still think there is going to be a cultural backlash to that and I think we better be very transparent and careful about how we go about it. But yeah, if it were up to me, I would absolutely turn my consumer purchasing data over to a health coach that was looking at my predictive risk scores. And if I knew it was going to improve my care, and especially lower my cost of care, I would absolutely contribute that data to that algorithm. I know that some people wouldn't but I think there are a lot of folks that would. There's a survey that came out this week that said most patients don't really care if you use their data for a research. You don't even have to ask. We always

	think they have had consent to do that. Most patients said they don't care. What they care more about is if you're using their data to benefit yourself from marketing. But if you're developing algorithms and things that will end up reflecting back on their improved care, most patients are comfortable sharing that.

[Dale Sanders]

Okay. Well we better shut down. Thanks everyone. I appreciate it. And thank you for the nice comments in the questions and things like that. And be sure and reach out and let us know what you thought about the webinar and other topics that we could talk about.

Tyler, back to you my friend.

[Tyler Morgan]

Alright. Thank you so much, Dale. Okay. Before we close the webinar, we do have one last poll question. Our webinars are meant to be educational about various aspects affecting our industry, particularly from a data warehousing and analytics perspective. We have had many requests, however, for more information about what Health Catalyst does, what our products are. If you are interested in the Health Catalyst introductory demo, please take the time to respond to the last poll question. Shortly after this webinar, you will receive an email with links to the recording of this webinar, the presentation slides, and any poll question results. Also, please look forward to the transcript notification that we will send you once it is ready.

On behalf of Dale Sanders, as well as the folks at Health Catalyst, thank you for joining us today. This webinar is now concluded.

[END OF TRANSCRIPT]