Paul T.: With the intersection of advanced data technologies and the evolution of healthcare, both providers and patients look to benefit. But security, privacy, and data hygiene still loom as potential areas of concern.

Hello, I'm your host Paul [Teese 00:00:28], and in this episode of If/When, [00:00:30] we discuss data science advances in the context of healthcare with two data science experts from Jacobs. Joining me for this episode are Dr. Jennifer Bloom, Senior Director Data Scientist, Jacobs CMS Cyber and Intelligence business unit, and David Morgareidge, Director of Predictive Analytics, Jacobs People and Places Solutions.

Well, Jennifer and David, thank you both so very much for joining me today. We're going to be talking about data science in the context [00:01:00] of healthcare management. And I know Jacobs is doing a lot of work in that realm, and also it's kind of a very interesting time we find ourselves in, obviously with the pandemic and whatnot.

So people are very focused on the healthcare experience, and so we're going to be talking about how data science can help improve that, both for the healthcare providers as well as for the patients. So I want to thank you both for joining me.

David, we're going to start with you. And [00:01:30] our first question out of the gate is, what do you see are some of the top benefits that prescriptive data analytics can provide to healthcare systems?

David Morgareid...: Well, let me start with a definition of what's foundational to prescriptive analytics. I think understanding what that means, making sure we've got a definition, is probably a good thing. And so prescriptive analytics is sort of the top tier of a three-tier evolution.

At the bottom level would be [00:02:00] descriptive analytics. So, simple tools analyzing historical data to understand generally what happened, how long did something take. Pretty basic.

Predictive analytics, which is my title in the firm, Director of Predictive Analytics, is a significant step up. You're still looking at historical data to understand what happened, but you've also got a simulation model, very detailed, that allows you to simulate the complete work environment [00:02:30] within which that historical data was created.

So in a healthcare context, you've got the architectural space, medical equipment, communication technologies, staffing models, patient arrival patterns, detailed process definitions, whereas you've got a faithful second by second representation of all of the work that produced all of that historical data.

And the predictive part is when you say, "Well, some of the KPIs historically weren't [00:03:00] the best. What could we do to improve those?" And so you make some changes to that model. You add another treatment space, you add two more doctors, you improve a clinical process, and then you run the model again. And you say, "Well, let's see, did that help move the needle on the KPI?"

And so I've done 50 of those, it's a major focus of Jacobs' work. But it's episodic, it's a project, it's got a start and a finish, and you're done. The key, the [00:03:30] big game changer for prescriptive analytics is the digital twin. What that does to this predictive model is it, 1) it lets the model take current state realtime data.

So there's data streaming from the electronic health record, from the realtime location to the service solution, and suddenly that model is in a current state, always, in a dashboard in the C-suites [backend 00:03:58].

Additionally, [00:04:00] it's got artificial intelligence - AI and machine learning - and now applications that are trying to look for trends, make relationships. And it's using that historical data to project forward an hour, two hours, four hours, eight hours, 16 hours. And it's looking for KPI problems.

So the institution doesn't want to have a wait time than an hour [inaudible 00:04:26]. It doesn't want to have the length of stay of a [inaudible 00:04:29] patient more than [00:04:30] 240 minutes. It doesn't want to have a door to doc time of more than an hour and 20 minutes. So as this digital twin is running, it's looking for, are any of those KPIs going to get compromised?

And if it finds a situation where it is, it then spawns another series of simulations that tries to look for answers. What could be done right now to prevent a KPI degradation eight hours from now? And it [00:05:00] produces a whole range of solutions and then offers those to the leadership of the emergency department or whatever clinic you're modeling, and lets them take action that avoids performance degradation.

So that's the prescriptive part. Predictive finds problems, but doesn't offer solutions. Prescriptive - in this digital twin environment - finds the problems and then gives you solutions to fix them.

So the benefit to the provider is that they've got a very [00:05:30] smooth-running operation. It helps them stay profitable, and it helps them deliver a good patient experience. It helps keep cost down so that the provider benefits and the patient benefits.

If you've got a newer car, is has active suspension system. It's sort of the same way... We used to have a car that had a shock absorber, a spring, and a tire. And you'd hit a pothole and you'd bounce all over the place. And these active systems [00:06:00] now have lots of data, lots of processors, and it's reading the

condition of the road a couple of thousand times per second and adjusting the operation of your suspension.

And that's kind of what these prescriptive analytic tools do in the healthcare environment, for a hospital or a clinic or an emergency department, or however it will be the technology is spread. In our practice, in the predictive model, we're doing one prescription [00:06:30] solution now for a confidential [inaudible 00:06:34] top-tier healthcare provider, and that's what we're allowed to say.

But in the predictive world, we've done projects that have saved clients in upfront costs: three to five million dollars, eight million dollars, a hundred million dollars. Over the course of the 12-year program, 983 million dollars. In ongoing annual operating costs, $782,000. So those are things that we know about in the [00:07:00] predictive project based model. And you can imagine that as this becomes a real-time effort, those same kinds of savings are going to be going on not over a project basis, but on an every-second basis. And the optimization concerns for the provider and the recipient of care, the patients, [inaudible 00:07:19].

Paul T.: Okay. And then on the flip side, how can data science help improve the quality of care that patients receive from their healthcare providers?

David Morgareid...: I think that [00:07:30] the key thing about data science is it's a way of dealing with very large data sets that are very complex. Human physiology and disease make a very complex overlay of a data problem. And data science, in the abstract sense, that's a right target for these kinds of tools.

If you look at what IBM did with Watson, that was the first artificially intelligence cognitive [00:08:00] computing environment that got a lot of public attention. It won a Jeopardy! game, it won a chess game. The very first commercial application of that was in healthcare. And the key about that is it allows personalized medicine to happen.

But I think that's... Just like in our design solutions, I did a project for the Army where I looked at 9,305 five different solutions [crosstalk 00:08:28] to find the best solution [00:08:30] for that particular environment. A lot of resources [inaudible 00:08:35]. 80,000 solutions in 36 hours, looking through lots of data to find the right solution for that problem.

And so the key thing that data science is going to do is that healthcare become personalized and targeted. So anybody who's had the TV on at least twice over the past two years has heard a lot of discussion about COVID.

[00:09:00] If you've got a five-year-old that's healthy, a 10-year-old that's got leukemia, a 30-year-old that's healthy, and a 65-year-old that's got dementia, those are not the same people. And it doesn't make sense to say there is a solution for people? Probably not.

And that's the key thing that data science has enabled us to do, is to find very targeted therapies that will provide much better care, [00:09:30] lower costs, more effectively than the mass, one-size-fits all solutions.

Paul T.: Gotcha. So we're going to kind of look at data privacy, of course, as a concern, especially with countries that have the GDPR in effect. And, Jennifer, I want to bring you in on this, and then, David, I also want to get your thoughts on data privacy as well.

But let me start with Jennifer. What are some [00:10:00] effective strategies you see coming into play that will balance potential data privacy concerns while allowing healthcare organizations and care professionals access to a patient's most sensitive data?

Jennifer Blum: Well, I think one of the most promising and cool technologies is actually the use of blockchain. So initially what you want is secure transmission of information. You want things to be encrypted, not just at the start and at the finish, but you also want it at every [00:10:30] stage, anytime it's viewed, anytime it gets transferred from one point to another.

And so you need verification and authentication that whoever is meant to receive or view this information is in fact doing so. And so that process can be really, really tedious, but the good thing is with blockchain, what you can do is, not only is that a cradle to the grave strategy for a person's information, but it involves [00:11:00] a person being able to make sure that they have their own private key, and a person who is going to view the information has their public key.

And there is that handshake, that exchange of information. And I think that would prevent a lot of things, and I think we'll probably talk about that later with preventing fraud and ransomware.

Paul T.: And then so, David, also, from where you sit, talking about data privacy, what are some effective strategies you see coming into play that will balance [00:11:30] data privacy concerns with allowing healthcare professionals to get access to the data?

David Morgareid...: I'm not sure what the right answer is to that question. I can tell you that it's... So I haven't run into the right solution, but I can tell you that the problem is pretty tremendous. In that digital twin project I was mentioning earlier, we have had to perform all work on the client's infrastructure, because we couldn't come up with a way to figure out [00:12:00] how to de-identify the data in a cost-effective manner that would allow us to have it on our network.

And in fact, some of the data that we determined was required, is not de-identifiable, so that's a real challenge. And there are some solutions that we're

looking into. We have a good relationship with Microsoft Azure, and the initial approach to using AI and ML for this [flight 00:12:29] was in the cloud.

[00:12:30] And they have since said that the process of de-identification is very complicated. And in fact, it's not as simple as saying, "This element is okay, and this element is not okay." Because you can have two elements on their own that are okay, and if you put them together, suddenly they're protected health information, and they're not okay. It's complicated.

And actually there's a third-party firm that has to validate that you've done it before it can be used. [00:13:00] So it's a really complicated process, and we haven't yet found a good solution within the context of this project. One thing I will say, though, is that just recently a small healthcare system in Massachusetts and New Hampshire have committed to move their entire electronic healthcare record and every other of the 300 data applications they have, to Amazon's cloud service.

And that's something that Amazon is broadcasting all over the place, [00:13:30] because it's a huge step forward. We don't know exactly what they did to make that okay from a data privacy perspective, but actually I'm going to try to contact them and see if that's something that we can leverage with this client that we have. So I haven't run into the solutions that you presented, but I've certainly seen the problem of what it means if you don't have a better solution [inaudible 00:13:55].

Paul T.: Yeah, no, I think it's fascinating. It's interesting because the [00:14:00] context can be very important. It's not just the information, but it's the context. And I'd add a conversation with a digital journalist, an author, and he talked about this back when the GDPR was just coming into effect. To that point of certain elements are innocent on their own, but when you combine it together, can...

So I think the example he gave, it's like you can tell foot traffic into a hospital a certain time of day, but [00:14:30] then you can also pair that with, say, somebody could find out when somebody got an Uber, took the Uber to that location, and then... And so you can start putting together this picture, "Oh, well, Jane Doe took an Uber at that time, and she ended up at that location, and the foot traffic was in the hospital."

And while it didn't identify Jane Doe, you can put two and two together and kind of figure out that Jane Doe was visiting the hospital. So [00:15:00] it's... Yeah, it's a bit more tricky than you might at first blush think, so...

Now, Jennifer, in a similar vein, can you speak to some of the potential cyber security pitfalls of an expanded digital footprint in the healthcare industry and how do you see those being overcome?

Jennifer Blum: Two of them actually that we're already seeing and that we'll continue to see as we get more digitized, is ransomware, which I think we're... It's in the news all the time unfortunately, [00:15:30] from banks to hospitals, where records are either locked, systems are shut down, until you pay the agreed amount to the hackers or whatever entity is doing it.

But also what you'll see is I imagine more events that are fraud-based is even more digitized. So if you're no longer required to come in physically to validate who you are, how would you do that? And so [00:16:00] one methodology that I think would be really useful for helping to overcome that - at least, starting to - is what I mentioned before with the thing of blockchain, which is secure encryption where you have your own private key and there's a public key available to whoever you want to view the data. And you both need to make that agreement before anything is shown.

I think that'll definitely help with ransomware because blockchain is multifaceted, so multilayered, and gets really complicated. And I love [00:16:30] it, but in a nutshell, it's that no digital trespasser can view what you see. Unlike, I think in a more traditional sense, where if someone clicked on a phishing email and all of a sudden you're in.

But if it got into the system, it didn't matter, because they still didn't have the credentials needed to even... They could see that there's data, but they have no idea what's on it, [00:17:00] so it doesn't matter. They wouldn't know necessarily what's important or what's not. You can make the assumption that it's all important, but I imagine it would be a bit stopping them a little bit more in their tracks.

And on the fraud side, mitigating that, would be... With blockchain, you can actually add to it. So if a person goes into a doctor's appointment, that gets categorized. Piece of data. And then they go back a month later. Piece of data. That gets tagged onto it. That is, those things [00:17:30] become immutable in a sense, and so you can't go back and change it for fraud purposes. So if you went last month, you can't say you went twice last month. So I imagine it would also save some people some money.

Paul T.: Hmm, interesting. And then adjacent to that is the topic of data management practices too. And, David, I think you and I have talked about this in the past, about poor data management and the impact [00:18:00] that it can have on projects and stuff. And clients want you to wave a magic wand and the data's just not, it's not what it is.

So, Jennifer, what are some of the best data management practices you're seeing or hearing about in the healthcare industry?

Jennifer Blum: Oh, so excited! So one of them actually is having a centralized system, so the electronic health record. Love it. Most people think of it as their patient portal.

You go in, assuming [00:18:30] you're assigned to a particular health organization, you go get your blood drawn or whatever, the records pop up there. Another doctor needs to look at it? Great, they can do so. It's one-stop shopping, and it's awesome.

I really like that I'm starting to see more of that now. The trouble that I see that a lot of places that aren't doing it, is because you just need to get everything into that system. So nurses' notes, any imagery [00:19:00] of say any MRIs or any just ad hoc notes that are put on paper, that all needs to go in there. And I think places are getting better with that, but I think it still has a long way to go.

And the second best practice to me is standardizing the information. So if everything's in a different format - and I don't just mean you do a PDF and I'm doing a word doc. I mean, literally.

So I'm based in Europe. They write the day [00:19:30] first and then the month. But in the States, we do it the reverse. Something as simple as that can cause great errors with a physician in a much more extreme sense, where it's medical data, because they're not quite sure what they're looking at.

Paul T.: And then my last question for both of you... And, David, I'll start with you. And then, Jennifer, I'll ask you, is...

So, David, what do you see as some of the more intriguing developments that machine learning is going to bring to healthcare data management in the next decade?

David Morgareid...: Let me go back to the cybersecurity thing just for a second, [00:20:00] because conversations about malware or ransomware typically focus on financial elements. And in healthcare, there's also a clinical component. And that one hasn't been discussed a whole lot, but [inaudible 00:20:16] University just recently did a study where they were trying to determine what impact degraded system performance, however that can vary. Either it's not working or the data's not available or it's slow or whatever the issue might [00:20:30] be.

How does that impact a clinician's ability to provide care to a patient? And they started with heart attacks. And their conclusion was that for every 10,000 heart attacks, there are an additional 36 deaths due to degraded system performance resulting from malware or ransomware. So I think there's a whole nother impetus, at least in the healthcare space, that's going to be added to figuring out [00:21:00] why this has to get fixed.

So back to the machine learning thing. The key thing about machine learning is that the data has to be there. And if the only data that's available is that that large systems have - like Kaiser Permanente or the Cleveland Clinic - it's going to short-circuit a lot of other smaller organizations from [00:21:30] getting involved in this process.

And so there's a market that's developed for firms that put together these kind of data packages. So Prognos Health is one of the firms that does that. They put together 500 different data sources. They've got 230 plus million people, 1 billion plus labs, and so this now becomes a data set against... If they've got the right folks in it, the data set that [00:22:00] you could acquire, if you wanted to research and you had your own little eight clinic system, but you were trying to compare your results against the larger pool. And that's the key thing that data science is about.

There are firms now that are trying to democratize access to that kind of data, which I think is essential for this to happen. The second thing is, well, once you've got the data, is it [inaudible 00:22:24]? I know in the work that I do, every single project that I've done in this space in the past 13 years - about 60 of them - [00:22:30] data integrity, data quality has been [diminished 00:22:35].

And there are now firms that are focused on that. There's one called [inaudible 00:22:42]. And they take these data sets from other folks, and they look for bias in how the data is assembled, they look for data integrity issues. And so if you're the person who's wanting to do research in this space, there are now [00:23:00] third-party organizations that are able to provide that kind of data to you.

I used Truven Health when I did the study once like this, and it was [inaudible 00:23:12], which has been acquired by IBM. And that was a phenomenal service. So I just preface the response by saying, the ability for smaller groups to have access to the kind of data that let them get involved in this is growing, and that's a good market.

But if you wanted to look for specific examples, [00:23:30] one example would be InnerEye. That's a project that's a joint venture between Microsoft and the United Kingdom's National Health Service. If you've ever looked at your own x-ray or your own CT or MRI or ultrasound, there's this fuzzy gray stuff up there. And somehow a radiologist is supposed to look at that and figure out, is there a problem? What's the problem? How do I treat it?

And in cancer, that's particularly difficult, to develop [00:24:00] radiation therapy programs depending upon what the cancer is, where it is, what's the right type of radiation therapy? Is it proton therapy? How many visits do you need? And what's... It's a whole, [inaudible 00:24:15] process.

And so this Inner ye program was an effort by Microsoft and the NHS to use artificial intelligence to help the radiologists. And radiologists are very highly paid folks, so it not only helps them to be clinically better, but it also helps drive [00:24:30] down the cost if you can reduce the amount of time that they need to spend on this kind of work.

And in this case, they've reduced the time to put together a radiation therapy program for a cancer patient by 13 times.

Paul T.: Wow.

David Morgareid...: So that's one teeny example of what's happening in that space. Drug therapy is another one. Pfizer has teamed up with IBM Watson to do immunotherapy research, again, against [00:25:00] cancer. So I think there isn't a malady you could find where there isn't somebody using data science in some way to try to better understand the [means 00:25:13] of data that...

An important point to me: a lot of times when you talk about data science, you're thinking about numbers. And numbers are important. But there's also content. And one of the first things that IBM did with Watson was to have it ingest [00:25:30] thousands and thousands of pages of clinical diagnoses and case studies and research.

Because part of what you're trying to do is not just let it look at numbers, but now what am I trying to find in that? And so there's an interesting blend of text, where there's intellectual content and data that represents a current state, and how AI can help look at the numbers smarter, [00:26:00] I guess.

Paul T.: And then, Jennifer, from your standpoint, what do you see are some of the more intriguing developments that machine learning is going to bring to healthcare data management in the next decade?

Jennifer Blum: So I'm going to piggyback on what David said first, which is with disease predictability. So I know he already discussed it a bit, but traditionally... This will be from a numbers perspective. Currently you use a risk calculator to try to predict, how likely [00:26:30] is this person to get this particular ailment? And you're at the mercy of knowing their routine, their genetics, their habits, all that stuff. But so you're at the mercy of just like six factors.

And they've found in multiple studies that they're just not incredibly accurate. But with machine learning, you can program a model or a simulation to build off of that, and it can ingest huge amounts of information. And I know that's something that Watson's [00:27:00] going to be definitely perfect for, but you can have up to 200 factors. That's all getting pulled from different devices, so - of course with the user's permission - you're wearing a smartwatch, and it's pulling in that information.

And the algorithms will spit out, for all intents and purposes, spit out a number. And it has a much higher degree of accuracy with trying to predict how likely you are to get a particular disease. So that's something I think that's really exciting. [00:27:30] Another thing that's less, I guess, healthcare management, just with healthcare in general - and David and I spoke of this - is with just machine learning, using it for AI and doing it for remote...

We're having remote surgeries from physicians, also having robotic assistance. So like David said, you have an expert looking at your MRI scan, and what are

you even looking at? But if a computer can tell you, "Hey, [00:28:00] these two things are a micron off," that's something that he or she wouldn't necessarily be able to see.

And then my own personal favorite with robots is, of course, the augmented lifestyle. So you have someone in a wheelchair, you can use AI to help them have a bit more of a robotic skeleton. And this is more than 10 years, but I'm super excited, but also they're currently working on bionic limbs. So [00:28:30] you have prosthetics that respond much more faster to heat or cold or what you want them to do, so that's more of improvement in lifestyle of the patient, less healthcare management. But I think they're both equally exciting.

Paul T.: Oh, very cool. Very cool. Well, Jennifer and David, I want to thank you both so very much for joining me today. It's really fascinating to see kind of where we're going in the healthcare sphere and where data science is leading the way. So [00:29:00] thank you both so much.