

Transcript for “Real World Analytics: Advancing Methods and Literacy in Healthcare,”  
Webinar  
June 23, 2020

Dale Sanders: Well, it's a sincere, heartfelt pleasure for me to spend time with Adam today. He and I go back many years. When I first met Adam, he was a youngster. He's still kind of a youngster relative to me, but it has been wonderful for me to watch his career grow because he's such a bright and capable and well-rounded human being in so many ways. So pleasure and thanks Adam, for joining. Let me read just a little bit of your background because it's impressive. You're a role model for professional development and your contributions to healthcare, friend. So I'm going to embarrass you with a little bit of your background. I didn't realize that you finished a double major in physics and math at the university of Utah. That was powerful, way to go friend.

He went on to study and get his PhD in medical informatics at Columbia. His advisor was a brilliant contributor, George Hripcsak, so great mentorship there. Then he went on and came back from Columbia back to Intermountain Healthcare and studied under another luminary in medical informatics with Paul Clayton. And Adam led the implementation of our primary care and emergency medicine systems while he was there.

He and I worked very closely on the development of the Intermountain data warehouse, which was pretty novel at the time. This is in the late 1990s. So he and I have been connected around analytics for a long time. And then he went on to be faculty and professor up at University of Washington where he's doing great things, not just at the University of Washington, but he's also doing a lot of things nationally with HRQ and the academic medical centers in the informatics space.

So what we get with Adam today is someone who is brilliantly smart academically, and also combines that with really boots on the ground, pragmatic healthcare operations. So it's an honor, and it's a lot of fun to spend time with you, Adam. Thanks for joining friend. I'll turn it over to you now.

Dr. Adam Wilcox: Thank you, Dale. Just testing, can everyone hear me? Okay, great.

So, huge thanks... I think that what I reflected on is Dale was introducing me, what I was probably most grateful for is the reflection on the people that I've worked with in my career. And that's probably that, and it's been a while since I've been called a youngster. But just reflecting on the people that I've worked with in my career, it's been an honor. And Dale is one of those that has been an honor to work with. And I feel like I've, in some degrees, been chasing kind of the ideas that he had back when we worked together at Intermountain for a long time. And maybe all this will be as a validation of that, but if nothing else, just a little bit of different perspective on it.

So I want to start, as mentioned I'm up in Seattle, we've made the news a lot recently in terms of autonomy and what's going on, but I want to go back a little

Transcript for “Real World Analytics: Advancing Methods and Literacy in Healthcare,”  
Webinar  
June 23, 2020

farther into our history to begin with, and kind of what created the civilizations around this area.

So what you're seeing here is an image of a Chinook salmon and early settlers in the area were drawn to the Puget Sound and the rivers around that because of the abundance of the salmon. That wasn't the only type of hunting that they did. On the point of Juan de Fuca at just the entry to the Puget sound, the Makah Native American tribes would actually build these long boats and go out and hunt whales. This is different than other whale hunting that happened in the rest of the Sound.

They were the only group who actually would build up boats and go harpooning. The others would hunt only beached whales. But I like this image because it kind of dramatizes that experience for them. This is the image of the evergreen huckleberry, which has also kind of a common food that was used by native tribes. And then this here is an image of a Honeycrisp apples and Washington state is very well known for their apples. In fact, Washington state produces 60% of all of the apples that are consumed in the United States. So at this point you might be thinking a couple of things. Number one, it's like looking at your watch, wondering when is lunch, it may be your lunch hour for some of you as I'm talking about food here, but probably also, you're wondering, why am I talking about this?

And so I want to use this as a good introduction to what I'm going to talk about because I walked through there a story from hunting to gathering, to farming. And that's... I want you to think about that analogy as I talk about these analytic methods, because in data and analytics in healthcare, we've really kind of transitioned through the first and second parts of that. When we started out, I remember when I first was working with Dale, before what we now call meaningful use, and it was at an institution that had already adopted electronic health records, but not fully and not everything, even though they were one of the leading organizations. And so if you want it to get data at that time, it was basically you eat what you kill. You have to go and figure out how to get those data collected in order to use it.

And now we're in what some have described the data tsunami. And even though we still may not have all of the data that we want, we have enough that as an health services researcher, or as an analyst in healthcare, you can easily find a place to gather up the existing data and study it. So we're now moved to gathering, this is a diagram... It's interesting, they kind of stopped charting this after a while from ONC in terms of the adoption of electronic health records. It's, I think, now about around 99%, but there's more data than we ever imagined that we'd be able to deal with. And that just happened over the course of one decade. So that's pretty impressive, that movement there. But that transition, what do we do with it as we've now gone to from hunting to gathering?

Transcript for “Real World Analytics: Advancing Methods and Literacy in Healthcare,”  
Webinar  
June 23, 2020

And you might think, well, it's a great time and maybe we should just stay here, stay there. Except that when you look at cultures, gathering societies actually aren't sustainable. They usually get pushed out by cultures that have figured out farming. So my goal is that to talk about that we can't stay in this gathering stage, that we have to find a way to get to farming. And there are different things that we talk about in, as an industry, about what can help with that, whether it's common data models and query tools and analytics applications. And those are things that, we know, can help get more people involved in going to be important to that. But broader than that are just a better understanding of the methods of using data. And that is going to be the conversation I'm going to have with you today about what are examples of these methods of using data?

Now, a lot of these you're going to have seen before, there's quite a few of them. I'm not saying that these are comprehensive, but I'm going to tell some stories about each of these in terms of how these methods are related to using data and analysis and the importance of them, because they're, each of them are kind of making make or break things. And in addition to that, just reflecting on how hard it is... These are some of these methods or some of the things that make analytics difficult. Now that doesn't mean it's impossible, it's just something that we need to be aware, what the challenges are. And I'm going to walk through those challenges with two acronyms. With COVID, which is one that we're all pretty familiar of right now, if we're not fully tired of it, but we still need to keep our masks on, and with LVAD or left ventricular assist devices, I'm going to walk you through an analysis that I worked on there, where it kind of shows the importance of some of these methods.

This is a paper that I did with it. It was a result of work that I did while I was working with Dale Sanders at Intermountain healthcare, when I was there for the first time. And it was described where we built, where we were transitioning from a paper, actually was a grease board, a whiteboard tracking system for the emergency department systems at Intermountain healthcare and moved it to this electronic version. It's since been replaced as they've migrated through, like everyone does, and to eventually commercial EHRs.

But I was part of the development and the design of this system here and also part of the evaluation. So we had this and we basically replicated what had been on the grease board, but added in the capability to link back to existing records. And when we did that, it was interesting because we found as we track the use of the system that allowed them to quickly see "Oh, someone has been here before, there are previous notes on this patient", that it actually was associated with a decrease in the rate of admission or the odds ratio, odds of admission from the emergency department, as we tracked that over time.

Now this is one hospital in one ED over one period of time. It's difficult to prove causality from that observational study or that quasi experiment, but it was still interesting to see. That's not what I really want to focus on more. I want to

Transcript for “Real World Analytics: Advancing Methods and Literacy in Healthcare,”  
Webinar  
June 23, 2020

reflect on what we saw is we worked with it with this diagram. So one of the physician... the content experts that we worked with, about a week after we had installed this system and we were pretty happy with it, because it was working, it wasn't crashing and it had done basically every function that they had asked for, I talked to him in a meeting about how was it going, and he looked at me and he said "I actually hate it".

Which surprised me because they'd been asking for something better and I just kind of observed that it seemed that it was easier and more consistent in collecting information. There were some significant improvements. So I kind of followed up and said "tell me what is it that's wrong with it?" And he said, "well, it's not so much the board itself that anything individually is wrong, but it used to be that I would come in to the ED and I could just quickly look at the tracking board and I could kind of tell what type of a day that was and now I'm not able to do that anymore. I have to interpret all of the information that's here to make sense of it in order to kind of move forward and I don't like that. It's, it's frustrating to me cause it's interrupting what I was able to do before".

And that helped me a lot thinking about, it's not just the elements that we have in terms of our information display, that actually the sum of the parts is much less than actually the whole thing. The display itself, people got used to seeing this is a really intense day in the ED versus this is not an intense day, just by the density of information and the colors that people were using. And so, as I reflected back on that, when we think about what we're doing for information display and dense information displays, there is great value, not just in putting as much information as possible in there in ways that people can easily assimilate, but actually showing them broader things about this is what it looks like in different contexts. And so they can assimilate that.

And so that's one of the important methods is how we consider information display and that idea about the whole is greater than the sum of the parts. Another example, and I'm going to go into talking more about context here. So this was a project that I led in New York city. It was called the "Washington Heights/Inwood Informatics Infrastructure for Community-Centered Comparative Effectiveness Research", which was a really long phrase that got us a good name called WICER. So it was the WICER study. And what we were doing was integrating a lot of data sources that we have from around the system, including going out and reaching out to the people in the community and serving them and request to link back to the data. Now, this doesn't seem so novel now, it seems to be done much more. All of us, as a program, is doing this at a national level, but this was back before any of those were done in kind of this precursor to a population health research database.

And when we were doing this, one of the things that we started to do that I thought I was most proud of the investigations that they did, especially Sue Bachen, and he later took over this grant, was the considering the information

Transcript for “Real World Analytics: Advancing Methods and Literacy in Healthcare,”  
Webinar  
June 23, 2020

and how we give it back to the community. And so we studied a lot in terms of how do you take this information and feed it back to people so that they can understand themselves in terms of context. And this is one diagram that was studied, where they looked at four different areas where you assess your health that are important for assessing health and how you might reflect on that. And it wasn't just showing where you were, was also showing what it was for other people. In some cases, that was the ideal. We have actually other examples where we weren't just showing the ideal, but we were showing people around you.

So sometimes it's not as helpful to show just the ideal because none of us are going with ideal. If you think about your experience in terms of being in social isolation, from COVID-19, a lot of the different things about health that may be useful to know, not just how you are relative to ideal, but how you are doing relative to other people around you, because there is... That baseline seems to change across the whole populations. And so this was really important for understanding, not just the data, but giving context to the data. Another important method is around sustainability. And I worked stemming off of that wiser grant and considering the data infrastructure we were building, what did we need to do in terms of considering sustainability? And so I was eventually a guest editor for a journal issue all around data and sustainability.

We had one paper in this were where we considered a lot of the issues around sustainability were one of the first components about building this sustainability plan is considering what your assets were. And in this project we had at first, we thought we considered in terms of the structural assets, specifically the data as the most valuable component. But as we work through analyzing those assets, and also looking at them over time, it was interesting to reflect on that the data themselves are oftentimes seen as the most important asset, but over time they become less important.

In this case, we had these data that were fully dependent on our interviews with the community. And if we stop those interviews, then within five years, those data became pretty stale. They would still be interesting, but they declined in relevance. The collaborations across the investigators and learning how to do this type of work and how to do outreach to the communities, et cetera, actually became much more sustainable and much more valuable.

And looking back at it, it actually turned out that the methods that we learned, both in terms of linking of the data, in terms of understanding how to display the data in context, but also in terms of linking with the data, understanding the data in context. But also in terms of understanding how to work with different data types and their relevance and the patterns of those data and how they're used turned out to be even much more sustainable. Data sustainability is a critical component and reflects even, not just the data, but also the other components strongly in methods, as it turns out. I wasn't involved in the work

Transcript for “Real World Analytics: Advancing Methods and Literacy in Healthcare,”  
Webinar  
June 23, 2020

directly here that was on data quality, but data quality is turned out to be a very significant method that we've all experienced. And I hope that we've all experienced it to some degree, if you haven't, that means you're not paying enough attention to your data as you're submitting it.

I have rarely seen a dataset that hasn't gone through some rigorous evaluation of its quality that isn't sent back or frustrating to the end user, because it takes a lot of understanding what was really done here. And so, when I was working with the Patient-Centered Outcomes Research Institute for Cory, with their methodology committee, as a member of that, and we were trying to establish what was known in the field about data quality. We met with people from this group who had built out this data quality assessments, specifically Michael Kahn, who was really significant leading this group. And he had commented on, at the time, back in... And this was just five years ago. We really didn't even have a good understanding at that time, have a good vocabulary, around what data quality meant.

People would talk about what we need to do in terms of understanding data quality. And we had a good idea of what it looked like when it wasn't good, but we didn't have a good vocabulary around what are the components of data quality. Eventually this group focused on specific areas regarding completeness and fidelity and plausibility in terms of their data. We actually took the recommendations from that paper and included them in the PCORI Methodology Report as one of the standards and telling people when you're working on an analysis. And actually this was important, it's not just the data set itself that needs to have quality analyzed, but for specific uses, you need to analyze the quality, how to look at conformance and how to look at other... Plausibility and data completeness regarding that.

And if possible, then how to perform sensitivity analysis to work in that. So that's another example in terms of these methods of data quality, and those of us who have done it, it's not like "Oh, you write a query, you get a report, the report populates some data it's done". You actually have to go back and figure out, does it look like it's supposed to, if someone is someone starts to interrogate this data, is it going to be recognized as having the right information? And that affects a lot of things. Now, just thinking about our own experience right now through COVID, we're seeing this a lot in terms of, well, what is the infection rate into different population? What we're shown in terms of our data is often reflected by what can be measured, not by what's actually out there.

And that's led to a lot of conversations in terms of how the numbers change according to our testing rate and issues with asymptomatic disease. That's just live and real tangible example of the importance of understanding data completeness and data qualities and example. So another method that I want to talk about is access or the, some people I've heard called this, the

Transcript for “Real World Analytics: Advancing Methods and Literacy in Healthcare,”  
Webinar  
June 23, 2020

democratization of data. Now, we all would like as many people as possible to have access to use data so they can do what we call data driven decisions. Or we talk in terms of learning health system where the information that we're able to gather from data is incorporated really easily into the decision making at the point of care.

Now, the challenge with that is that we talk about that as a good idea and yet we've got a significant barrier to getting there just in terms of do people have access to data in a usable way? So I built this diagram of this funnel kind of showing what the issue was and try to make it really tangible, just in terms of how long does it take you to actually get to the data that you need. And there may be many people on this webinar who have spent a lot of time accessing data and working with it, and maybe like riding a bike, you forget how hard it was to learn to ride a bike after, well, after you've been riding for some time. In some sense, getting access to data or getting understanding of the data can be something that we forget how hard it is for other people to then follow in a similar pattern. Why is it taking them so long to get the data? Well, once they get it, it seems like everything solves, but I want to reflect on that first part: how long it takes someone to get actual access to data. On the left side, I'm going through the different types of people that can get data and on the right side, we're going to talk about how long to get. I'm going to focus on how long, and then kind of reflect on what that means for who can do it.

If it takes a minute to get data, and you think about what that looks like, that's data, dashboards, displays, information, reports that people have built up, where it gives the information in an easily accessible way, that can take about a minute and everyone can have access to that data. These are raw displays or dashboards that we give a lot of access to, people are able to use that pretty effectively.

If it takes an hour, we've stopped being able to use it for what I would call point of care clinical use. The patient, if especially an outpatient, the patient's probably already gone by then. If it's an emergency, the emergency is being addressed differently. If it's making a decision about who I should see next, and it takes you an hour to get there, you've lost enough time to make that not the relevant decision anymore. Usually people, if it takes an hour to get the data, have to be doing things not at the individual interaction with patient level, they're doing things more at a system level. Leaders, for example, will be able to take an hour gathering data from whether it's from different sources or in complex ways in order to use that. If it takes a day to get the data that you need, even leaders often don't have time to do that. And so there has to be another level of interest in that specific area. Experts may often take the time to get data if it takes up to a day.

Now, when you get to a week or a month, the group of people who can access data when it takes a week or a month to get access to that data are generally

Transcript for “Real World Analytics: Advancing Methods and Literacy in Healthcare,”  
Webinar  
June 23, 2020

people who have in their job description something about getting data; where that is their job. Whether it's a week and they're getting data generally, or whether it's a month and they're getting data specifically for developing for a specific task, these tend to be much more limiting in terms of who can get access.

All of us when we reflect on, Okay, who is it that actually has access to our data, what does it take in terms of them getting there and all of the permissions will reflect that we'll generally have a group of analysts or developers who have access, and then a lot of people who are making requests through them. Because otherwise it's just so difficult to get there. Sometimes, actually, in some situations, the process of requesting the data may even take at the longer end of this time.

Now, that's where we are. What do we do about it? As I've reflected, internally at both different organizations I've been in, but also in my current organization, we do have instances where it can take a month to get data. Usually that's because either people don't know where the sources are or they don't know what the processes are in terms of requesting access. They'll spend some period of time spinning their wheels, waiting for responses on emails from different people in terms of access requesting and determining permissions.

One of the first things that can help that is just instruction. What we call our analytics curriculum, helping people understand what data we have, how to get access to it, what requesting and how it can be used. And that can generally, it doesn't eliminate it to down to in a minute, but it can speed things up. Whereas before it may have taken a month, now it can be done in a week because you, at least, are steering people in the right direction to begin with.

If it's taking a week to get data, oftentimes that's because you're trying to find where the data are and the process of going through that and having data coming from different sources and you're having to then build, write those queries yourself and interpret all of the different variables and different columns that turn out that they're significant in terms of what they represent, whether ... I mean, we've all gone through this where, "Oh, I didn't realize that that was the column for a canceled visit versus all of the visits." Something like that. What can really speed that up is actually, if you build out a query library. Different people have learned how to do this, and sometimes it's this tribal knowledge that we work with with smaller groups. But if we're really going to make this more democratic and make the access much more broadly, that tribal approach doesn't work as well and we need to actually build out structures such as this query library to support it.

Moving from a day, if it's taking you about a day to get data that's often because just navigating the data structures, figuring out how the data are actually represented, what it means to be a visit, what table those are in, which is

Transcript for “Real World Analytics: Advancing Methods and Literacy in Healthcare,”  
Webinar  
June 23, 2020

sometimes not very obvious, and so I'm a big proponent of actual data modeling to improve that where you're going to build out tables. We've moved to a point actually where you don't need to redefine a model. We have existing common models that one could use. But if you put the data into those models, people who would before take up to a day, which is a lot of that work was just understanding how the different tables interrelated, can move it into an hour. I think that these are really important methods, both the analytics curriculum, building query library structures, and data modeling. But their importance is most demonstrated in terms of what does that do for who can actually now access and use the data effectively.

Dale Sanders: I have to say, buddy, I take a sort of pride and maybe arrogance in believing that I've thought about just nothing is new to me in this phase by now, but that's the most creative way and framework that I've seen to represent this. Your physics background shines through. This is great work.

Dr. Adam Wilcox: Thank you, Dale. This is probably the most original thing that I actually have on here so I actually appreciate that. But this was reflecting kind of in the trenches experience, trying to solve a problem and reflecting on what would actually do that. That's kind of how we build something. What's ironic about this, as you're saying that right, is I was going to this time move into talking a little bit more about maturity models. I took those out because I'm not going to talk to Dale Sanders' group about maturity models given that Dale Sanders has had almost a decade, well, more than a decade, of work in analytics and maturity models. I can tell about my experience in using them, but that would be much less original and so the contrast of you saying that is actually pretty funny.

Dale Sanders: Very nice.

Dr. Adam Wilcox: I started, when Dale and I worked together was actually at Intermountain Healthcare in Salt Lake City. I'm putting this up for a couple of reasons. You might think, well, I started out by talking about food; now, I'm going to talk about scarcity. That's one way to think about it. Another way to think about it is that this is the ultimate example of shelter in place in history.

But there's actually a deeper message that I'm going to give here about the Donner Party so bear with me if you can for a bit. In 1846, the Donner Party left from Independence, Missouri, which is where a lot of settler groups traveling to the West started from and moved, trying to get to California. As they went through, they did a couple of different things. One was to take a shortcut and they waited for Hastings to come and tell them how to do that shortcut. While they were waiting, they built a road somewhere between Little Sandy River and the Great Salt Lake Desert, going through the Rocky Mountains there, which later on turned out to be extremely valuable for one year later, for the Mormon pioneers, as they went much of that same trail, at least getting into Salt Lake city.

Transcript for “Real World Analytics: Advancing Methods and Literacy in Healthcare,”  
Webinar  
June 23, 2020

But while it was helpful to the Mormon pioneers, it was more significant to the Donner party because they'd spent some days building that road and waiting with instructions for that cutoff and then kind of proceeded across, got across going through the Sierra Nevada, and we're just one day from sumitting in the Sierra Nevadas when the snows fell. It's this tragedy and because of the depth of the snow of that first snowfall, that they had to effectively shelter in place and the rest of it, I'll leave you to look up on Wikipedia in terms of what happened if you don't have a lot of background knowledge on this. But it's a pretty tragic story of the American West in terms of what those people had to do to survive.

Eventually, they were rescued and they got out. One of the survivors, Virginia Reed, later in a letter to someone describing something said, "Remember, never take no cut-offs and hurry along as fast as you can." I'm making this pretty dramatic because, well, we're working in healthcare and analytics. The stakes are not quite the same as they were to the Donner party, or the Donner-Reed party as it sometimes called, but reflecting on what that means for how that affects analytics and our knowledge management.

I had the opportunity to work at two organizations who were exceptional in terms of their knowledge management. That was in and around healthcare where Stan Huff had done a great work in building out the Health Data Dictionary, and Columbia University, where Jim Cimino had built out the Medical Entities dictionary. These were two knowledge management structures or vocabularies or medical ontologies, however you want to call them, as I studied as a student, these were the examples that I worked with a lot and a lot of these definitions of them were used.

But it was really a different time then, because again, that was before in the hunting stage when you kill what you eat. At that time, when we built those for knowledge management, there was a different use of them. At that time, we talked about how data were modeled and we talked about deep concepts and surface forms and the hierarchies of the data and how they were represented. Perhaps, most importantly, the decision support needs were actually driving the documentation approaches. So if we think about some of the initial requirements as part of meaningful use, one of the most significant ones was the requirement for physician order entry.

Why was that? What was the big change that led to people saying we need position order entry? That was heavily influenced by the idea that there were errors that were made in terms of prescribing medications, dosing or drug allergy, drug interaction where, if we could get the people, making those decisions to do it electronically, we could do decision support to test against those. That was a driver for this new model of documentation. Actually, the continuance of that approach has led to other challenges that we have faced,

Transcript for “Real World Analytics: Advancing Methods and Literacy in Healthcare,”  
Webinar  
June 23, 2020

leading to actual significance in terms of physician burnout and how EHRs are being used.

But that was then. If any of you have been part of an EHR adoption process where they're building out that database or they're building out the system, there's not as much time that gets spent in terms of building out the knowledge management as things are going through. We all know when that happens that what's on time and on or under budget is the measurable metric that people are tracking; in terms of that, and whether or not you take the time to build out the vocabularies as you're building them, it seems to be less important. While building out the vocabularies and the knowledge management structures may be helpful to the people actually doing it, it could be seen like building a road when you're the Donner party that you're trying to get things done as fast as you can and trying to not miss the deadline by one day. And so you're not going to spend a lot of time with that. That's significant now that we've got through this widespread adoption, but also significant in terms of the utility of that knowledge management and how it's done.

Now, we talk much more in terms of how data are stored and data definitions and phenotypes becomes much more important. Whereas before we were thinking about ways we can model the data as it was going in, we're now looking at phenotypes as example of what are the breadcrumbs about what's really being represented from different data sources even. Rather than decision support driving documentation, decision support now is much more likely to be based on predictive analytics and pattern recognition approaches.

I cannot overstate the significance of this transition in terms of the informatics field and in terms of its influence on analytics. In fact, I still see in the field, many of the leaders still grew up in that then stage, and I will acknowledge that not that I'm a leader, but I'm one of that population that grew up in the previous stage, in the hunting stage, where we were worried about how data were modeled, and we were worried about these hierarchies, et cetera. It takes effort and a change in focus and thinking in order to move to the use of data as driving the knowledge management rather than the storage of it.

I went through those methods, information display, data context, sustainability, data quality, data access, and knowledge management. Again, not to be comprehensive about what was going on in terms of these methods, but to give an example of these complex methods, there are actually challenges in the field that we have to deal with. You may be in conversations with people about why is data and analytics in healthcare difficult? What are some of the reasons? These methods may be some examples that you can use for that. Again, I think we've all experienced our own methods that may or may not fit into these examples I gave directly, but hopefully they've at least added to the list of things to consider.

Transcript for “Real World Analytics: Advancing Methods and Literacy in Healthcare,”  
Webinar  
June 23, 2020

I now want to, as I mentioned at the beginning, go through two examples. The first one I'm going to go through is around COVID, but I'm starting with an image of the covers of two reports that I was involved with back when what is now the National Academy of Medicine was called the Institute of Medicine. About eight years ago, for a few years, I was on a committee that was working on advancing care for people living with HIV and AIDS. Part of that was understanding monitoring the care in the United States, looking at strategies and what were the data systems that would be required for that. Again, this was shortly after meaningful use. We hadn't moved fully to a gathering stage yet, and we were reflecting on what is possible with the HRS in terms of monitoring this.

Also, I think it's important to recognize, this was looking at HIV and the data systems that can be used to monitor. COVID is the first big pandemic where we've actually passed this move from the hunting stage to the gathering stage, where we have widespread adoption of EHRs and the potential to use those data. When you think about in that context, it can be useful to reflect on, well, how are we doing with that? With HIV, when we were looking at it, we use an important model, which was the Engagement in the HIV care cascade, which is a good way to reflect kind of what data you need at different stages for monitoring people. You have a population of people that are infected, and then of that, there is a smaller set that actually know that they are infected. They've been diagnosed.

With HIV, each of these stages, it feels like it's about 20% decline across them. So another 20% drop off between those who are diagnosed versus those who are actually linked to care, and then those who retained in HIV care, and those who actually have appropriate therapies, and then eventually those who have the demonstration of the outcomes, which is a suppressed viral load. This is what we used in terms of monitoring the data.

If you think about where we are with COVID, there actually is a similar treatment cascade. The percentages in terms of the drops are different in terms of the number of people infected versus those who actually know they're infected is pretty significant. It's much more than 20%. When we look at the strategy that we had with HIV, those that we did not know that were infected, those that were not yet diagnosed but may have been infected, that 20% was significant enough to drive public health outreach of wearing condoms. You could reflect on, well, that was a 20% drop here. We have at some estimates, an 80% drop between those who may have COVID versus those who actually tested positive. Diagnosed as having COVID with an 80% drop, what is the approach that you do?

At first, it was shelter in place and extreme social isolation. Now, but the other responses and kind of thinking similarly around HIV in the same way that public health is pushing heavily for use of condoms, broadly, the use of masks would

Transcript for “Real World Analytics: Advancing Methods and Literacy in Healthcare,”  
Webinar  
June 23, 2020

be a similar prophylactic or barrier approach because you're not positive of who would be infected versus who was actually tested positive for it.

Dale Sanders: Adam, I just want to comment on that slide. If you can go back. One of the concerns I have with COVID is that I don't see at a national level, a framework of thinking as reflected in this slide. You can cut and paste HIV and replace it with COVID, and this is how we should be thinking about COVID, and it might be happening at institutions, but at the federal level, I don't see the evidence that it is. I think that's just worth noting and I'm going to bring that up. In fact, I'm going to use this slide in conversations related to this at the federal level. Thanks for bringing this up. It's a great way to look at it.

Dr. Adam Wilcox: No problem. Well, what I like about this is this is a really good example of a data driven ... the model of a care cascade is not unique to HIV as I mentioned. The data across this are really informing the approach.

In the later stages, in terms of what it means to be diagnosed and what it means to be cared for appropriately, and it may be with COVID since there's varying approaches, but one of the most important one is, because the spread is so significant, the equivalent to link to care, maybe that you're diagnosed and actually that you're informed about what you need to do for social distancing and quarantining. Self-isolation, and quarantining, and that those around you, maybe, in what you do. Then, also, that you're actually monitoring and getting the appropriate care and escalation of care as you need.

It would be useful to kind of go through that exercise because I agree. I don't think we know what the numbers are. We have vague ideas about that first step. That it's around an 80% drop, which is huge. It may be similar like how many people that are tested positive are actually self-isolating? It feels like we have so many examples of people who, these super spreader events where one person that was sick that decided to still go out in public, and what they're linked to. There seems to be some gap there. Even if it's not, you're tested positive, that's symptomatic, the issues of just even being symptomatic may be a way to reflect on that.

Dale Sanders: And with that kind of a framework, you really have a different data strategy for each of those steps in the cascade.

Dr. Adam Wilcox: Totally agree.

Dale Sanders: I mean, there's some overlap, but there's a lot of differences in the data that you're collecting on the left side of the screen, which is public health data, versus the right side, which is real world data, real world evidence or outcomes.

Dr. Adam Wilcox: I cannot overstate how important this diagram was to these reports because when we reflected on what were the data systems that we needed, it was

Transcript for “Real World Analytics: Advancing Methods and Literacy in Healthcare,”  
Webinar  
June 23, 2020

specifically based on: well, what do you need at each stage? The idea that just one would cross all of them was really flawed. It took a very, like kind of a surgical, a very specific approach to it.

Early on with COVID, I'm in Seattle, we are not the greatest hotspot currently. We've been surpassed in being a hotspot by lots of States, but we were the first one. Initially when it started, I remember looking at our data and well, other people may have had a hundred, we quickly had 4,000 cases, and was wondering what can we do in terms of providing this data out for other people to analyze. I actually even reached out to a colleague at the CDC saying, "We've got these data. What can I do to be helpful so that other people can learn from our experience as we're going forward?" That was back early March. I did not anticipate it spreading like this where everyone else would have experienced it at much drastic ways. I don't think anyone anticipated what was going to happen in New York until we started seeing the data.

But I got back from the CDC this as a request to complete this table. I'm just going to kind of reflect on what they were really asking for here and what's one of the examples of what makes analytics hard. This was for each week, can you provide for us, and this is just for an understanding of where the people being cared for and stuff, the people in ambulatory care, urgent, ED, in the hospital and in the ICU. Information on the totals, the number that are tested, the number of positive so we get an idea of the penetration for how widespread is this and across these different age groups. That's a lot of data. What's interesting is it's less data. At the time we may have had 4,000 tests and about that point, there were about 10% positive. So 400 positive instances, and they wanted this by week. And so by this time we'd gone through a couple of weeks, two or three. So, and then they also requested, can you get this for, different race and ethnicity as well.

So that if you calculate that, there's five categories, three each across, going down. So that's 15 times seven age categories is 105, multiplied by however you want to divide out the race and ethnicity. We quickly have more queries or more cells in this table. Then we actually had people positive for the disease and we were the hotspot at the time. It's not that way anymore again, but still, I don't know who can provide this type of data reasonably on call to organizations.

And so just because we had the data didn't mean that we could make it useful because this reflects, well, in order to use it, this is what they had to do. This was a real eye opener. Oh my gosh, you just don't... if we're going to be providing data for you, like with these reports, we're not going to be able to do it. What would it take to give you access to it? And could we do that, and we really haven't reached that point.

Transcript for “Real World Analytics: Advancing Methods and Literacy in Healthcare,”  
Webinar  
June 23, 2020

In fact, I've been participating in what's called the National COVID Cohort Collaborative or N3C. This is not the only COVID sharing initiative out there, but it's a significant one currently. And it started with some work on a grant that I was one of the investigators on with, which was CD2H, which was funded by National Center for Advancement of Translational Science, or NCATS, where we were trying to build out like a structure for people to share data. Academic medical centers, can you share together information from COVID?

And as we were talking about this, initially I had advocated strongly for this because one of the points I made is I feel like we're at the point... We've spent a decade now after meaningful... where we went through meaningful use, a decade post the high-tech rules, where we were worried about the sharing of information and the breaches. We've spent a decade where analytics is general of healthcare data has generally been performed by analysts within those organizations,

And a decade into this, I can look at it and say, I don't think we're going as fast as we need to for this. I think like we're limited by the number of analysts that we have within those organizations. And I experienced this within COVID. Like we have really good analysts at the University of Washington, but we don't have so many that we can be answering every question. Wouldn't it be nice if we could open this up in some way and make it a safe way to share these data, at least for COVID, with other people who can then bring their analytics skills to bear on this and actually help us all. And so that was kind of one of the motivators for this initiative. And so we started working on, okay, let's share data.

So I had an analyst working on our team who had finished, just finished his PhD, but more important than that before that he had worked in government. And so he was pretty good at kind of documenting what the steps were in reset. Like, as we tried to share data, can you document what it took for us to actually get to the point where we shared data? And so we went through this process and bear in mind, this is in a pandemic when people are recognizing the importance of doing this, and we've got national support for this, of what it actually takes to share data.

So on the left side, building the data out into a common model, that was actually the easiest part about this. The approvals in terms of research and the IRB determination. And then we used a centralized IRB in that caused some more like how that integration to the process caused some delays, and then just getting the organization to agree, because there are organizational privacy issues that need to be addressed, even if you can do it in the limited data set and have it secure enough, there are some of those issues to be addressed.

And then actually the process of finalizing those organizations or organizational agreements and doing it, like there's a lot of steps in here. So it was really

Transcript for “Real World Analytics: Advancing Methods and Literacy in Healthcare,”  
Webinar  
June 23, 2020

important for me to kind of see that just because we can share data doesn't mean we actually are able to. Just because it's possible, doesn't mean it's facilitated and doesn't mean that you can do it in what would be considered an emergency and a desire to do it quickly.

And it was kind of shocking to realize the difficulty of doing that and those two things about the need for sharing data, but our ability to share data. The need for sharing data so that organizations who need it, aren't dependent on analysts querying out, as I showed here, a thousand data elements for that. But they could have access directly if it were in a common model and the need for actually sharing this so we can pull data together and understanding together. It was disappointing on both points about what it would really take. And I think that this is one of those areas where we're going to need to do better, and we're going to need to take the lessons learned from here.

So the last of all, let me just go through this example and I'll go through it quickly in terms of what we discovered. So that was my COVID story. And now let me talk a little bit more about a left ventricular assist devices. So these are for people that are pretty sick. It's where you have a device that comes in and helps their heart move blood around. These are historically high risk populations. If you're looking at the Kaplan-Meier curve in the upper right corner, it shows that after five years, about 50% of the people have died from this. Causes of death are usually related to bleeding or not bleeding or clotting in some way.

So most common cause of death is stroke. The most common adverse event, for those that don't die, is internal bleeding or other two within the top five are stroke or re-exploration for bleeding. So a lot of the issues around this use of LVADs are going to be effected by bleeding. There's two ways that they can kind of monitor the bleeding. The most commonly used model is where someone uses a partial thromboplastin time or PTT, which measures the time it takes for blood clots to form. And we were looking in our organization and we should really be using a different measure, which is called heparin anti 10a or factor 10a or 10a. I'll use either of those to describe it, which measures the amount of heparin by its inhibition of factor 10a activity. And you can see the therapeutic ranges here that are used.

So the first thing we did is we said, well, we actually had a lot of data. We had thousands of data points. Actually I think it was about over 6,000 data points where we had taken both PTT and anti-10a measurements on individuals. And we first wanted to see, well, did they always match up? And if they didn't, what was the issue? And so we first built out this discordance or discrepancy matrix, and you can see from this, that in general, there's that 47%, which seems to be pretty high on where factor 10a is listed as medium, but PTT is listed at high. When we actually plotted out the various points on here, we saw that yes, that

Transcript for “Real World Analytics: Advancing Methods and Literacy in Healthcare,”  
Webinar  
June 23, 2020

was the case, that middle right cell was pretty packed there. And as we looked into it in more detail, we kind of looked at some issues around the data bias.

So PTT, again, the therapeutic range there was 60 to 100. And you're seeing there that it's kind of biased high, whereas factor 10a, it's biased low. And the spikes, those of you who've worked with lab data, know that that's what happens when you fill in everything that's left less than 0.01 or greater than 200, what you do with those data points. We put them all up to a one or then 0.09.

And then we looked at, okay, well, what's going on? Are these representing the same thing? They're not quite, but, well, let's look at these distributions. There wasn't much that we could see in terms of the distributions between a PTT value frequency and those who bled of differences versus factor 10a, it didn't really reveal that one was better than the other in terms of separating out the, like having a bi-modal distribution.

We also looked at regression lines for that when we plotted out the points. And while there was a difference that factor 10a was better at predicting bleeding. No one's going to change healthcare based on a... regression lines on data like this, that's just... there's too much uncertainty there. And then we even did ROC curve analysis. Finally, actually what we did is we had to model the clinical decision, which is a threshold decision where we looked at it and said, okay. So if someone has... If we choose a threshold at some level, whether it's PTT or factor 10a, what's the probability of lead above that level. And so we modeled that threshold decision, which Stephen Pauker in the 80s kind of talked about the threshold decision making of medicine. And, but we hadn't actually defaulted this.

What's notable is everything until this point, they're kind of the standard analyses that one would do and what you actually, most of the machine learning or statistical analysis would do. When we did it this way, we actually noticed an immediate separation between them, where factor 10a was, there was actually a clear distinction that at higher levels, it was able to detect a higher bleeding, whereas PTT really didn't do so much.

And so we actually used, this is the chart that we use to change how we were using these different lab results for monitoring bleeding for these patients in our hospital. Now, the most interesting story though, is actually how we got here. So what had happened was we had in our system eight bad outcomes in a row. And when you have eight bad outcomes in a row and given, the bad outcomes for this, there's a lot of negative outcomes in terms of bleeding or bleeding events or stroke events or death.

So if you think about that as a 50% odds, what's the probability of flipping a coin eight times and coming up heads? Well, it's one over 256, you'd say that, oh, of

Transcript for “Real World Analytics: Advancing Methods and Literacy in Healthcare,”  
Webinar  
June 23, 2020

course, we've got to investigate. Something is wrong here. And the doctors started to investigate those eight cases. And one of the things they came up with was there seems to be this difference between factor 10a being more predictive than PTT. And then we went and took the data, 6,000 plus measures, and all of this analysis just to demonstrate what they found out at first.

So all we were doing was giving the data behind what they had already detected with eight cases. How would that happen? What is it about people and their ability to look at data? And more importantly, what is with this odds? So as it turns out, that coin flip probability that I said was one over 256, it's actually not that. We had over a period of few years, over 200 instances of LVAD use.

And so if we're looking for what's the probability, when you have a run of eight across 200, that's much more frequent. It turns out it's not just one over 256. You got about a one in three chance of seeing over that time sometime where you're going to have a one of eight. So it's because of that, that they stopped. And then they looked at it. And then, because people are really good at identifying patterns, they're not really good at interpreting trends, but they're really good at identifying patterns.

They picked up the pattern that it took us all of that analytics work to figure out. And I've reflected on that a lot saying, what is it that we could do better with data? Because we could have found eight negative instances long before we waited for it to be eight at once, eight in a row. We could have looked at, showing people, having them do chart biopsies on eight negative examples. And they probably would have noticed this pattern much earlier. And we probably would have been able to improve a lot better. And what can we do? And kind of the consideration that maybe what leads to improvement is being smart about how we use the data to identify the right clusters, that then allow those who are able to see the patterns best to see them. So I'm at this point going to stop and turn it over to Brooke.

Brooke MacCourtney: Great. Thanks Adam. So we know it is the top of the hour. We will stay on for a few more minutes. We can ask some questions. If you have time to stay on, please submit your questions on the go to webinar panel right now. I do want to mention our HAS summit before we jump into our Q and A. Our health analytics summit, it will be held virtually this year, September 1st to the 3rd. And our theme is on the role of data and analytics in our new normal with COVID. And we'll be featuring speakers who battled COVID in the trenches, as well as other speakers who are adjusting and pivoting to this new normal. And we plan to provide a unique and innovative virtual experience. We're going to have some nationally recognized keynote speakers, a few of which you can see on this slide. We'll be announcing a few more soon, and we'll also have some individual connection points with our analytics walk about with our networking, with our brain date and other virtual activities.

Transcript for “Real World Analytics: Advancing Methods and Literacy in Healthcare,”  
Webinar  
June 23, 2020

We're really excited for what we have planned and we hope you'll be able to join us. Registration is now open. You can register at [hassummit.com](https://hassummit.com) and it's \$99 per registrant. And right now we want to give away three complimentary passes to HAS. So if you know that you're going to be able to attend and you're interested in being considered for one of these three passes, please answer this poll question that I am just about to launch. So I'll leave that open for just a minute for everyone to answer. And let's give you a few seconds to answer, and I've got one more poll question, and then we'll jump into the Q and A. Few more seconds. And thanks to everyone for staying on and listening to what we have to offer today.

Dale Sanders: Yeah. I can stay. Adam, can you stay for a little while?

Dr. Adam Wilcox: Yeah, I can stay for a few minutes. Yeah.

Dale Sanders: Okay, great.

Brooke MacCourtney: And also for anyone that needs to jump off, we are recording this. So if you need to leave and you miss the Q and A session, you can listen to it later. We will include that. All right, I'm going to close this poll and we've got one more poll question for you. So while today's webinar is focused on real-world analytics and advancing methods and literacy in healthcare, some of you may want to learn more about the work that Health Catalyst is doing in the space, or maybe you'd just like to learn about our other products and professional services. So if you'd like to learn more, please answer this poll question. I'm going to go ahead and leave that open while we jump into the Q and A. So Dale, I'll let you look at the questions we have coming in and kind of see if we can answer as many of those as we can.

Dale Sanders: All right. Thanks Brooke. Thanks Adam. Really interesting stuff there, friend, I've got a bunch of questions and there's several from our audience here. I want to go back to the pattern recognition that humans can sense these patterns, and you were essentially validating those patterns with the data around the LVAD patients, right?

Dr. Adam Wilcox: Yeah.

Dale Sanders: I love that, friend. I agree with that. And this'll sound cliché, but the older I get, the better I am at pattern recognition. You just kind of build up expertise. I wonder if we should, as analysts and data people in healthcare, I wonder if we should occasionally make room in our day for having conversations with the frontline care workers about the patterns that they're seeing, that they're probably stuffing away in the back of their mind, but they don't always verbalize. But if they were prompted, it might lead us to identify patterns and then investigate with data to see whether those patterns are valid or not. What

Transcript for “Real World Analytics: Advancing Methods and Literacy in Healthcare,”  
Webinar  
June 23, 2020

do you think about that? How do we get to the pattern recognition Part of the human brain as data analysts.

Dr. Adam Wilcox: Yeah, so let me give you an example of that, and this is actually a credit... So this happened the second time I was at Intermountain Healthcare. I was working with a primary care clinical program. And it was based on the work, Dale, actually, that you had done in terms of there and building up the warehouse. So I was meeting with a group and they were going through the diabetes, the indicators. And so they were really good at kind of building up incentives and helping people track their patients that were meeting certain criteria of high quality care, but not everyone met them.

And in this meeting, one of the doctors said, I actually took that report and I went, and I pulled out from that, all of the individuals that I had were not meeting these metrics. And then I went and looked into their chart about this, and I noticed something. And so what he noticed is actually less important than the process. So think about what he had to do. He had to take a report. He had to effectively run that same report or the query. I even know whether, at the time we didn't have the equivalent of slicer, dicer, some other query tool within the EHR, but he had to run that query to gather those individuals. Choose a set of them, browse into their record. And just by looking at it, like 10 of them, was quickly able to identify some pattern that wasn't necessarily in the data.

So I think that that's going to be perhaps the most important part, those chart biopsies and enabling those chart biopsies. What is it that we can do? Now, if it's just us reviewing the chart and understanding what's documented in there, if it's the clinicians, getting them to be able to see that information.

Because I think there's probably... when I reflect on what that doctor was talking about, some of those elements, I think, were things that actually weren't documented in the chart, but he just knew the patients and started to see those patterns. So the more that we can do to get the... what I like to call is, subpopulation analysis. And to make it easy for people to drill in to those instances within there, I think that's going to be what's going to generate the greatest knowledge generation, those hypotheses that then we can use data to determine. Because that's actually all we did with the LVAD is that there was a hypothesis we just validated it.

Dale Sanders: Yeah. Right. So maybe the lesson for those of us in the analytics world is that we should be a little more proactive about conversations with, especially, and I mean this very sincerely, especially with the folks who've been around for a while, right? Because you build up pattern recognition the longer you're in an environment. And ask them, are they seeing any patterns? And then from that generate hypotheses that you can test with analytics. I love that concept. I think you're really onto something there. I love the term chart biopsy too.

Transcript for “Real World Analytics: Advancing Methods and Literacy in Healthcare,”  
Webinar  
June 23, 2020

Dr. Adam Wilcox: That one I did not make up.

Dale Sanders: Yeah, right, right.

Dr. Adam Wilcox: Subpopulation analysis and chart biopsies. Those two, one is analytics-driven and the other is the drill down. And there are scientific methods to do each of those. I actually think that is the... Much more right now. And, and I may, I hope I don't offend people by saying this, but much more important than machine learning, much more important than predictive analytics.

Dale Sanders: Yeah, I think so too, friend. I think you're onto something there. I want to figure out how to propagate that concept in the industry. Speaking of chart biopsies, what's a data phenotype?

Dr. Adam Wilcox: So a good example would be, and you can think of this, because I remember when Steve Barlow and you, when I first met you guys and you were just building out the data mart for diabetes. So what does it mean that someone has diabetes? They have diabetes because someone documented that they had the condition as a diagnosis code. They can have it because their hemoglobin A1c is about some value, or they can have it because they're being treated with Metformin or other medications that you use for treating diabetes.

So the phenotype, the way I best describe it, and you don't have a formal definition, but with that is, it's a definition of the combination of those different data sources that indicate the broader finding, which is patient has diabetes. That's the phenotype. And the phenotype definition is, in that case, as you notice, isn't something that can be defined just by a vocabulary because it's crossing three different data sources. It's a more complex analysis of what it means to have diabetes.

Dale Sanders: Yeah. I like that concept. It's, what's the expression of the disease in data.

Dr. Adam Wilcox: Yeah.

Dale Sanders: Yeah. I like that. That's very interesting, which by the way, has been quite a challenge with COVID. I'm thinking I need to, and maybe you can help, the notion and the importance of a registry, like a COVID registry.

Dr. Adam Wilcox: Yes.

Dale Sanders: And recognizing the subtleties of that data phenotype in that COVID registry. I'm seeing across the country, there's a need for sort of registry continuing medical education or something. I feel a lot of struggles around just understanding the concept of a COVID registry and then the subtleties of its design. And I like your term data phenotype.

Transcript for “Real World Analytics: Advancing Methods and Literacy in Healthcare,”  
Webinar  
June 23, 2020

- Dr. Adam Wilcox: And it's not just for COVID. I mean, in places where they've had to expand ICU capacity, it's no longer that they're on some floor that they're in the ICU, it's what treatments they're getting. And other things like that.
- Dale Sanders: Yeah. There was a question from Rob Naaman in the audience, back to your time to get data slide, is there a timeframe between the minute and hour that keeps it in the point of care model? The timeframe between the minute and hour, so tens of minutes or something?
- Dr. Adam Wilcox: It depends on the setting. If it's a 15 minute primary care visit, the doctor can't spend those full 15 minutes trying to dig something up. And so you think about how long, and you reflect on the information of the info button, how long will a doctor spend digging down to try and answer a question rather than focusing on more gathering of information or other interactions with the patient. And I don't know that we've figured that, but we know it's not 15 minutes in that case. Now with others, I've talked to doctors, especially hospitalists who may spend longer trying to figure out a difficult patient and stuff. And so that kind of can because you know, they're residents or the attending's in the hall.
- Dr. Adam Wilcox: ... the residents or attendings in the hospital setting that can be a little longer and outpatient care could be limited. So I don't know, but I think it's on the lower end of-
- Dale Sanders: Yeah.
- Dr. Adam Wilcox: ... closer to a minute than an hour.
- Dale Sanders: Yeah, me too. I think you've made it closer to seconds. I would say that there's a timeframe below minute, one second. Yeah.
- Dr. Adam Wilcox: Right.
- Dale Sanders: Now back switching gears here, going to the HIV cascade slide. How long did it take to settle in on those values and sort of work through that cascade, both the development of the framework as well as the data required?
- Dr. Adam Wilcox: There's a question which is how long did it take experts in the field to settle in on the treatment cascade as a way of defining that?
- Dale Sanders: Yeah.
- Dr. Adam Wilcox: And I don't know, when I was working on that with that group, I wasn't the HIV expert, I was an informatics expert who had joined in with it. So I didn't have a lot of detail for that. I think that HIV in terms of treatment and filling out the rest of that cascade and diagnosis of it had been around for a couple of decades. I think it probably, as the people were trying to figure out what it meant, looked

Transcript for “Real World Analytics: Advancing Methods and Literacy in Healthcare,”  
Webinar  
June 23, 2020

a lot more like... Not confusion, but the lack of information across the whole thing that we have with COVID was probably similar with HIV as they were pursuing different methods and trying to understand it.

But you're going to try the people that you're treating and helping the people that you're seeing directly, you're going to figure out from public health how to prevent it from coming in. The cascade became a really good way of reflecting on what are the issues and of organizing around the data. And I think that because of that, it was probably... I'm going to guess it was probably a few years into the pandemic there that they came up with that, but I don't know the full history. I just know it was really, really useful when we were identifying what are the data needs and those data sources.

Dale Sanders: And yeah, the treatment cascades sort of trend. You could also in some way, call it or not cascade treatment cascade, but disease cascade, the transitions of care is also kind of a similar concept to measuring the movement of a patient through its disease state is, is an interesting observation for me and COVID. And so I'm going to throw out some comments here and I'd love to hear you react to these. So EHR data content has been influenced by ONC, right? So interoperability and the mission of ONC is influenced what we collected at EHR. It's been influenced by CMS and payers, and those motivations have been around quality measures and reimbursements, and those motives are reflected in the data that we're collecting into the HR. And so I would argue that the today's EHR is sort of, the content is dominated by the motives coming from interoperability quality measures and reimbursements. And what I see in COVID very starkly is a lack of adequate real world data and evidence to pursue cascades of care, understanding, and also outcomes analysis. And so think about that for a second. And then I want to also advocate that I think someone federally, if not federally (silence).

Dr. Adam Wilcox: Did I lose you Dale?

Brooke MacCourtney: Yeah Dale, we're not hearing you. Don't know if you went on mute on accident?

Dr. Adam Wilcox: So hoping he comes back on, let me reflect on trying to answer where I think he was going with his question.

Dale Sanders: What do think about that, and also comment on OMOP and Odyssey.

Dr. Adam Wilcox: I mean, we're in the middle of COVID and when I look at what we're trying to do with N3C, if you were to build a registry with CA with HIV care, it's much more easily defined because we have enough experience with the condition that we kind of know what data elements to track. And one of the things that you didn't mention that drives care is also, or, I mean, that drives data that we have, is also a clear understanding of what data are important. And I don't think we've reached the full point now yet with COVID. I mean, there are questions, is it the

Transcript for “Real World Analytics: Advancing Methods and Literacy in Healthcare,”  
Webinar  
June 23, 2020

blood type that's most, that actually leads to it, a lot of different hypotheses are flying. If we were to build a registry with, again, with HIV, we kind of know most of the elements.

With COVID, I think there's probably more that we don't know in terms of data elements that we would need. We could have a pretty good idea of the first hundred, but, or maybe the first, probably the first thirty, but after that, there's going to be debates on whether that's really important or not. And I'm just reflecting on, we didn't realize that race and ethnicity information was going to be more important than so many other things at first, and we're still learning a lot. So I think it's kind of the maturity of an understanding of what data elements are most important, and there are building influences, but we're still not sure of what building influences that have changed. A lot of the quality metrics that have been documented have been because of definitions of what quality care looks like.

I don't think we know yet what quality care looks like for COVID. And that's one of the reasons to want to use a common data model, like OMOP, where you can grab a whole bunch as you figure this out. So you have flexibilities that rather than saying, we don't know enough yet to build a registry, we can say, well, we know enough to start. And if we get the data models as well, then maybe we can abstract from that as we go. I mean, if you think about those 300 plus elements that the CDC was asking for, they weren't quite sure that all of those were needed either, but we hadn't defined kind of best practices of data.

Dale Sanders:

Yeah. I would have loved to have seen at the federal level, a convening of the major EHR vendors and in that, in those meetings and iterative process around defining the minimum viable data sets that we needed for outcomes analysis around COVID from the beginning, realizing that it's going to be chaotic and nebulous in the beginning, but lay down a minimum viable dataset that the EHR vendors then would take out to their deployments and help implement. And then that flows up, we learn from it, we iterate, we improve. Now that also assumes that the EHR vendors have the ability to quickly turn that thing around, which generally speaking, they don't, because of the architectures and things like that. But I think that's where we've got to start thinking as a country. And we've got to recognize that the real world data that we have right now, isn't as great as we think it is and someone needs to own that space.

Dr. Adam Wilcox:

Right? Well, so actually, credit to the EHR vendors and the federal institutions, I think they actually did that to some degree, but not as widespread. And I don't think they did a convening across them, but I know that they did work with Epic, and I think Cerner, directly in terms of the querying of those data that I showed, I think that they went to the EHR vendors as a place that was at least convening multiple sources of data, rather than asking each institution to kind of track that. I think they were pretty innovative in that way. Now I will argue that that doesn't solve all the problem because that still limits the analysis down to then

Transcript for “Real World Analytics: Advancing Methods and Literacy in Healthcare,”  
Webinar  
June 23, 2020

the EHR vendor's analytics. Cause they're the only ones who have access to all of it. And I think that's a big problem and that's something that we need to solve, and the fact that they didn't, and I don't know, it may have been because there was concern of people found out they were doing that, they may have complained about the use of data and stuff like that.

And in the fog of war, the fog of the pandemic, a lot of decisions, we all see within our organization, things that we chose to do that may be, if we had known what we know now we would have chosen differently. I think they have tried to do that, but even if they've been successful about that, I don't think that's the final solution, or I don't think that's necessarily a complete solution because I think that it's a more efficient solution than waiting for everyone else to figure out how to share data and my diagram kind of showed that, but I would hope that we get better at sharing data. So that if there's something else like this, and there will be, maybe not like an infectious disease pandemic, but something of national interest where data sharing is important, there will be other things. And hopefully we can figure that out.

Dale Sanders: Yeah. And I'll give credit to the MITRE-Mayo coalition for serving as a convening body of sorts. So I was part of an initial group of Health Catalyst, Cerner, Epic and Allscripts, where we got together under that concept of a minimum viable data set. But we never quite got there in terms of changing what data was being collected formally in the EHR. And of course the MITRE-Mayo coalition is kind of all about Goodwill and a little bit distant from actual care delivery. So I'm still, I'm still biased towards, I think someone federally needs to own this space and I think FDA might be the place to do it. Then we'll see how that goes.

Dr. Adam Wilcox: Yeah. FDA or ONC, yeah.

Dale Sanders: There's a question here from Carol Coates. I don't know if Carol still on, but we've got 87 people on. We will be implementing Epic in 2021. Do you have any guidance on reports that provide just-in-time analytics by patient population that we can put on provider dashboards?

Dr. Adam Wilcox: So, only medicine, we will be going live on Epic. We have an outpatient, but full inpatient in 2021 as well. So I don't know that yet. We're still figuring that out. I'm at your stage in terms of us wondering that Carol and trying to figure that out. I would defer to people who have more experience with Epic directly on that. I believe that's probably a good, I don't know if they have those webinars, exactly, but the lessons learned of what you need to know after, what are the real things that make a difference in terms of success? Not just determined by the EHR vendors as they're coming in, but by people who have gone through that meeting together and convening in saying that, and there's probably those groups going on, we just are not yet to the other side of it, but I would hope that the information from those would be useful for those coming afterwards.

Transcript for “Real World Analytics: Advancing Methods and Literacy in Healthcare,”  
Webinar  
June 23, 2020

Dale Sanders: Yeah. And I would all answer it very high level that the important thing is to think less about all of the possible analytics use scenarios that you might need and more about the infrastructure of data that can support lots of use cases, no matter what comes up. And one of the problems that we see as an analytics vendor and also in COVID is lack of standard terminology. So it gets back to Jim Smino and Stan Huff, right? And the lack of standard terminology is a real problem in COVID. And it continues to be a real problem. The DHR is to support reusable and consistent analytics. So for instance, lab harmonization is, in general, pretty bad across the industry. We've implemented lab harmonization is sort of a testing or preparing for the tests around meaningful use and that kind of thing. But beyond those core requirements for reporting lab values are not well harmonized in the country. And that limits the analytic use cases you can support. So spend time on the data strategy and involve, as you deploy Epic, involve informed analytics folks in the deployment, so that the design and the configuration of Epic is informed from the backend as much as it is from the front end.

Okay. Let's see here. Let's ask another question here. There's a couple of questions I don't understand, unfortunately. And so I won't go there. Let's go to this one from Brett Johnson provider pattern recognition in the weekly mortality morbidity stats, where are you able to build out survival trees from the groups with age? Did you start to see real time leading indicator patterns from pharmacy, EEG, steroids with outcomes?

Dr. Adam Wilcox: That type of analysis, no. And I mean, we built out some data displays internally that the ICU docs and the residents were looking at and trying to understand kind of what that survival and just for that, for them, it was these questions that we'd like to know what we're doing and what eventually happens to these individuals, those analyses we haven't done. I think that they're probably better done at places where there they've spiked more. I mean, University of Washington would be a good place just to study the spread and how we kind of monitored the spread. New York and New Jersey would be good places to study kind of in a ICU treatment. We just didn't have as many people intubated in the ICUs. We had a lot of people, I think our total inpatients was around 400 so far. I mean, we're still tracking it, but UW medicine.

Dale Sanders: Yeah. Okay. Let's go. One more question for we're down to 72 people. You've been around analytics forever, when we were at Intermountain way back when there were no vendors in the data warehousing space. Now there are dozens including the EHR vendors and the cloud has made homegrown data warehouses appear to be easier. So what are your thoughts about this space and in the role the vendors are playing like Health Catalyst, EHR vendors, Homegrown. What are your thoughts and advice to the attendees about how to approach an analytic strategy from kind of a technology vendor build your own space?

Transcript for “Real World Analytics: Advancing Methods and Literacy in Healthcare,”  
Webinar  
June 23, 2020

Dr. Adam Wilcox: Yeah. So that's a good question. And if I had included the information on maturity models, I would have pointed out what I think, and Dale, you're the one who built out the healthcare analytics adoption model. The first thing that usually throws things out, that makes things difficult, is the ability to centralize data and where if someone has a homegrown data warehouse, even within an institution that has a regular data warehouse and they've built something homegrown, it usually starts because, well, they have this data source that they needed to also include that they couldn't put into the warehouse because they didn't have quite the institutional mandate or the funds to do so. But they ended up instead kind of building a localized database where they were centralizing things that they could query it, them together, like that usually starts it, but it also starts eventually people saying, well, we need these other data.

And so it pushes them towards the centralized database for the institution as well. But if they're one offs, it's usually cause they're, cause they're gathering the data together, or they're trying to gather, to get all of the data in the same place where easy inquiry I think that, that's pretty, that's really significant. The design initial design of warehousing often is indicated most by what were the efforts around central again, around gathering data in one place? So the problem is, is that can either become a friend or a foe to level two, which is actually organizing the data. And I've seen within our own organization, we have activities that are centralizing data, but we have also other people that are centralizing with, they're localizing the data.

And often what leads to that is kind of the local use of the data, because perhaps there's not a flex flexible way for them to define those phenotypes or the information about the use of the data. I think that we underestimate the significance of the data organization and the importance of sharing that information. It's not that different. Like even at the simplest level, I would at least start by figuring out how to, how to organize query libraries together in an effective way that you can update them and that people can query them and sometimes vocabulary is going to represent them. But, and then all the next figuring out if you're going to be building your local copy of it, why wouldn't you use a common data model? Sometimes it's hard to put it in that way, but you're eventually going to be modeling data anyway. So I'd go back to that funnel diagram that if you can follow those things, you're actually going to get greater benefit at the end. So, does that answer your question? I mean, actually, I'd like you to reflect on it because all I'm doing is quoting your maturity model.

Dale Sanders: Well, it's an interesting space to be in, as a vendor. On the one hand, certainly the cloud has made homegrown data warehouses easier. I used to be well known for knowing how to engineer the infrastructure of data warehouse. Now, I don't really care about that. The cloud has taken care of that engineering infrastructure for us in ways that I never could in the past. The challenge is in the layers above the infrastructure now, in scaling a homegrown data

Transcript for “Real World Analytics: Advancing Methods and Literacy in Healthcare,”  
Webinar  
June 23, 2020

warehouse is pretty darn hard. It's pretty expensive. The EHR vendors traditionally come from a background is not about data warehousing analytics it's about data capture. So it'll be interesting to see how they evolve and catch up to that space. And yeah, so there are good options, probably worth a separate webinar to talk about how to make that decision.

Dr. Adam Wilcox: Okay.

Dale Sanders: All right friend, we have to jet, I really appreciate you, Adam. And I thank everyone that hung on and keep doing good work out there, everybody. Thank you, Adam.

Dr. Adam Wilcox: Thanks so much, it was wonderful to talk with you.

Dale Sanders: Likewise, thank you Brooke.