Intro: When you take a magnifying glass and look at a picture, you see that it's made up of thousands of dots and when you pull back, those dots become the details that create the picture. There's real power in the details, because when you have more detail, the bigger picture becomes sharper and wider, and a story emerges.

Hi, this is Amie Moreno and you're listening to "Seeing the Big Picture: Conversations on how Data and Artificial Intelligence can Add the Details that Fuel Deeper Insights in the Life Sciences Industry."

Amie Moreno: Hello everyone, it's Amie Moreno and I'm here to bring you another electronic health records related podcast. I'm the Director of Data, Advanced Analytics and Tools on the Optum Life Sciences team and, as a reminder, we are here in the Boston office which is conveniently located downtown, so while it's fun for us, we realize it can get noisy, so don't be alarmed if you happen to hear an ambulance drive by or any construction. With that said, today we're really lucky to have with us Shelli Field, who is going to discuss, again, the EHR data, particularly as it relates to commercial research.

Shelli comes to us with an extensive experience in a variety of market strategy roles. So, Shelli, thanks for joining us today.

Shelli Field: Thanks, Amie, I'm pleased to be here.

Amie: Great. Would you tell us a little bit about your role at Optum as well as your background?

Shelli: Absolutely. I've been in marketing strategy, marketing research, new product development for the bulk of my career, which spans over 25 years and we'll leave it at 25 years. I've worked in marketing research, both qualitative, qualitative meaning focus groups, in-depth interviews -- you know, you don't do them without a broad number, maybe 30 people, maybe 60 people, but that's about it -- and quantitative primary market research, so surveys would be the quantitative with a larger sample size. But in the primary world, a large sample size can be 200 or 400 and then, of course, I've been working with big data for multiple industries, including pharmaceuticals, and I had my own company for a little over 10 years. I've been at Optum for over six years at this point and when I joined Optum, Optum only had one primary data asset, which was the claims data, and I was brought in to help try and optimize our claims data custom work to see if we could integrate some primary work with that, not that we wanted to be a primary market research supplier, because that was not the goal, but to be able to look at the different data assets that our clients were using and how could we supplement and maybe bring that qualitative

perspective, the softer issues around that claims data to provide deeper insights to our clients.

Shortly after joining Optum, Humedica was acquired and then the role all changed at that point. I moved more to a commercial analytics consultant role and I've stayed in that type of role ever since. I've always worked with brand teams on the commercialization side, so I feel that that's where I bring the most value and insight.

Amie: Great. So, you mentioned your previous work in primary research, as well as the claims data, so in your role today in the commercial side, what other types of data do you use or what data assets do you use in conjunction with each other, for example?

Shelli: Okay, perfect. Well, the types of projects that I do in commercial analytics run from very in-depth patient journey work to pricing studies, forecasting, predictive modeling, kind of runs the gambit, lots of patient profiling, including profiling physicians based upon behavioral patterns. So, if you kind of look at the product life cycle, you can tell the different types of insight that are needed along that life cycle and the different data sources that we use typically are going to address issues along that life cycle. So, in the short-run here, in my work, I utilize a variety of data assets, including claims data, including EHR data, and then different linked data assets, not only including maybe some of the data assets we have in-house, but also with data that’s provided by clients.

So, in addition to working with various big data assets, we are doing some supplemental work in bringing qualitative, in-depth interviews, either speaking directly to physicians whose claims data is available, so I have their history, I know what they've been doing, they've been part of the analytics that were provided for the brand teams. We go back, and we interview them to get more of an in-depth look at, not so much what they're treating, maybe reasons more in-depth why, but also into, are there certain protocols within their offices that are helping dictate why they may be having these behavioral changes? So, that's been really interesting because it's a totally different use of the qualitative methodology than we typically would have used before. We also can use some quantitative findings to build recruiting criteria and then combining these methodologies really helps to provide, not just a broader perspective, but a more in-depth perspective as well.

Amie: So, like you mentioned, each data source is different, not every data source is the right one for a specific type of research. So, can you tell us some types of research or business questions that can be answered and how you would decide which data asset is the most appropriate?

Shelli: Sure. The choice of data asset is always dependent on the needs of the client. So, for example, if pricing or cost is of interest, then I have to use claims data, because claims data includes pricing information and cost information, it's based on billings, it's based upon whether it's reimbursed, so that information along with diagnosis and treatment and that kind of information is available. Claims data's particular value when you want to understand pricing impact of a new class of drugs or a new drug that's coming into the market. So, in commercial analytics, I've done quite a bit of investigative work in various biosimilars, sometimes using analogous drugs as comparators because the new biosimilars haven't been out long enough for them to show up in the claims data, but we can look at analogous drugs and see what those patterns are. The other thing that where we would use claims data is for target lists, and that isn't necessarily always used by the brand team itself, but more by the sales team, because we can provide identifiable physician-level target list. When I say identifiable, I'm really talking about identifiable, its name, its address, its NPI number, but we can add, or I have added additional profile and variables, so how many patients of a certain type is this physician treating? What does the treatment look like? Are they prescribing more of this or more of that? If we've done a project that is based upon doing some segmentation work, we can identify which segment from a prescribing behavior standpoint that physician belongs to or that their patients belong to.

So, all of this really allows the salespeople to totally target, who do I go to? What do I know about them? What do I know about their practice? And what's going to be the best way for me to deliver my message?

So, if the client's interested in understanding patients or physician behavior from a clinical perspective, I utilize the EHR data. The EHR data provides insight not only into lab tests that have been ordered, but also into the actual test results themselves. I've found utilizing this information, along with the richness of the descriptive information found in the physician notes or as I sometimes refer to it as the physician commentary, it provides me the opportunity to utilize multi-varied approaches to segmentation and predictive modeling; it's actually based on lab results, which can give a total different perspective when you're thinking about how do I frame the patient type in the mind of the physician? Now you're giving them something concrete to deal with when they're looking at that physician and they go, "Oh, this HbA1c is this, then, oh, product A comes to mind." Right? So, that helps with that. So, it really does help provide the client with very specific profiling insight, so they can help to develop their overall strategy, positioning, messaging, and so on. The EHR data has also provided deeper insight into understanding drivers for treatment and that can be symptoms, it could be lab tests, it could be other co-morbidities, or to determine if potential market opportunities exist among a

very specific patient segment. Because of the variables within the clinical data, we can totally hone in on that specific patient that is going to be appropriate for that brand team's product, where we can't do that with other data sources. So, that's the beauty of working with EHR data -- the clinical specificity it provides.

Now, an integrated data set is the best of both worlds. If you're just talking about claims and EHR data, that can give me insight into what the physician ordered, what was actually dispensed, or in the case of testing, what was actually completed along with the depth of treatment, physician notes, lab results, and then cost as well. So, you're really getting a panoramic view of what's going on within that whole treatment regimen, treatment paradigm, cost burden perspective, it just provides such in-depth insight that, you know, clients love it and we love working with it. So, we have the ability to understand not only like what has happened after a script was ordered, you know, was it filled? Wasn't it filled? Was it refused? Was it replaced? But then we can carry that forward, so along that step, those steps of the continuum, we can profile the patients based on the script behavior, which provides insight you can't get any other way. And then the ability to add the pricing and help your burden analytics, again, clients get a much deeper perspective of the issues at hand.

So, think about how the brand team uses this information. They need to develop the optimal strategy, they need to develop the optimal positioning, and they need to develop the most influential messages for their brands. They need to create those most influential messages to either healthcare professionals, patients, or payers, and to make sure that their brand is in the provider or patient or payers consideration set -- that's their goal. So, where's the best point of intervention? You know, at what point along that treatment regimen, along that patient journey, is the best point to have your message hit the physician, the healthcare provider, or the patient, so that they're thinking about your brand? Those are all things that we can find in EHR data. The EHR data helps us look at not only when, but where, why, and how. It really provides the foundation for their strategic platform.

Amie: Thank you for that. You provided a lot of information about the different types of analyses that you're able to do and you touched on why EHR is important, but given your experience, how has EHR enhanced the ability to do some of those types of analyses that you might not be able to do without having that asset?

Shelli: Sure. Well, working with EHR data is always a fascinating experience and particularly coming with a stronger primary research background, I was used to using both qualitative and quantitative data, did a ton of market landscape positioning, messaging, new product development work, as I

mentioned before. And while I'm still a fan of primary research, I've come to look at it from a new perspective. The EHR data provides me the opportunity to look into the treatment of hundreds of thousands, if not tens of thousands of patients and not just look at what their treatment has been, but have insight into the true drivers for treatment selection. I mean, after all, we're looking at the exact same data that the healthcare professional uses when they're making the treatment decisions. So, the EHR data provides me with the opportunity to see those qualitative aspects of a patient visit that healthcare professional records on the patient chart. So, I'm not just looking at, this is what they took, this is what they had, this is what was dispensed, but you almost can relive that patient visit because of what's being recorded within those patient notes. So, descriptors relating to things like symptoms, frequency of symptoms or severity of symptoms, what were their vital signs? What were there different lab results? In some, but not in all cases, the healthcare professional has recorded some of the over-the-counter medications that a patient may be taking.

All of these insights that you won't find as part of claims data or any other data source. So, that totally is enhancing the work that we can do. So, I always tell my clients that if I had had the types of analytics provided in EHR data -- and don't forget that huge sample sizes that are associated with these analytics -- the primary work that I used to do would have taken on a whole new meaning, it would have been more targeted, it would have had more specific questions, it would have been at a higher level of understanding and a better use of my funding, to be quite frank.

Amie: And I think that's an important point because I know you touched on this, to say, we're not trying to replace primary research.

Shelli: Oh, absolutely not, absolutely not. And I'm dead serious when I tell people that, you know, patient journey work in particular, should be the foundation work that everybody does, whether they are going into a new market -- I've done patient journeys for clients that aren't going into a market for five years, but they want to understand, what are these patients going through now? What are the treatment regimens? What are the drivers for treatment? As well as things that are already out there, you know, what's going on in the marketplace and how has that changed? The nice thing about both the claims data, any of the big data that I work with, is that I have the ability to look forward and back, so if I'm looking at a particular index event, we call it. Let's say it's a diagnosis of a particular disease, I can look back to see what led up to that diagnosis, I can look forward and see what happened after, you know, so, were they treated before? Or what was going on? As I stated, the data that I'm currently using, allows me to look forward and look back, so you have a particular index of that.

So, for one client, we were interested in type 2 diabetes and we were interested in those people who had type 2 diabetes who became insulin-required. We also had a control group of people that met the same requirements, an HbA1c was the main catalyst for this. We used an HbA1c of 7 and looked both from the point that the patients had an HbA1c of 7, we looked backwards to see what was their pathway coming up to that, and then going forward, the group split. Those who had an HbA1c of 7 who were put on insulin, those who were HbA1c who were not put on insulin. So, what that allowed us to do was to be able to look at what was that pathway coming up to the point that you are considered out of control? Most of these patients really ended up being at an HbA1c of 8 and above, which is definitely out of control, and then be able to look at, were there differences in treatment, differences in lab results, differences in HbA1c's and BMI's, over time, between those that were using insulin and those that were not. In this case, the client identified a market opportunity for a drug that they have coming out in a few years, hopefully, among this patient base. So, we couldn't do that anywhere else. We can't use claims data, we can't HbA1c, we can't get BMI, we couldn't get symptoms, we couldn't get all of that rich clinical information that the healthcare professional utilizes when they're treating their patient. So, I think that's a pretty good example.

Amie: So, earlier you mentioned sample size with the EHR data. Can you tell me a little bit more about why that's important in comparison to other types of data that you've worked with in the past?

Shelli: Sure. Well, sample size drives everything. It drives the reliability of the interpretation of the data. Brand teams are not as specific about their statistical requirements as like HEOR. If HEOR is doing work, they're looking at P-values all over the place. Branding team, they really are interested in 95% confidence level. When you're dealing with the sizes of sample that we have, which is usually thousands, tens of thousands, hundreds of thousands, in some cases, millions, it's all statistically significant. So, if you see a shift and maybe it's only a 1% in shift, it's a true shift, you know, it's statistically a true shift. It's not, gee, at the 95% level, you have to have a, you know, 12% difference for that shift to be statistically true. These are accurate, statistically relevant findings and there's no place else to get that kind of numbers behind what you're seeing from treatment regimens.

The other thing, let me throw in, which I just thought about was when I first started doing this, and I was working with claims data at that point, and so I knew the physicians that I had done work with, I knew what they had been treating. In the primary world, when you're looking for treatment, you'd say, "Oh okay, their next 100 patients, or next 10 patients," whatever, you know, divide up how you would treat this. So, I actually had

what was filled, I was working off claims data, so I knew what had been prescribed because they had filled for it. We did some qualitative interviews with some of these physicians and what they said they had been prescribing or whatever, really wasn't even close. And I was shocked. I mean, I really was shocked. So, the reliability that you're getting when you're actually dealing with whether it's the claims data or whether it's the EHR data, because that's the actual script that was written, is far superior to what you're going to get in any kind of a recall data.

Amie: Right, and then when you build in the intersection of the claims, then you've got the full, complete picture from physician intent to patient behavior.

Shelli: Yes, exactly.

Amie: So, what happens when you build in an integrated data, for example? So, whether it be an EHR database linked to a claims database, for example, or even if you bring in other novel types of data?

Shelli: Well, the more data -- it's not always the better -- but in a lot of cases, the more data, the better. And so, clients continually use multiple data sources every day to provide insight into their business issues, and they don't all have to be data sources that are amongst the exact same patients or things like that. But when you get that data source that actually has the same patients, that's like the Holy Grail of things, because you're looking at what I have access to is I can utilize a deterministic link, which is based upon name, address, social security number, you know, it's not information I see, but it's information the people that actually do the data linkage see, so the data I get, I'm able to look at the integration of both what happened within the physician offices from the patient charts standpoint and then actually what happened from a dispensed standpoint along with that.

And let me give you an example why that's important. Many times, we can see if we're just using claims data, we know that a test occurred. If we're looking at EHR data, we know that the physician said, go have this test. We don't always know if they actually went and had the test within the time that they're supposed to be having it. So, when you have the integrated data, you can actually see what they were ordered to do, and did they follow through or not, because if you look at something and you say, well, they were supposed to go have the glucose tolerance test or whatever and then the next time you see them and you're looking what their HbA1c is and you're like, "Wow, what happened here?" You find out they never went and had that, that's an important thing for a brand team to understand, that being compliant is not a strong suit among a certain patient segment or whatever. So, that kind of information, I think, is always

useful. And the more data linkages that you can get -- so some clients would like to have more lifestyle information, so can we build that information into that patient profile? So, you've got active versus sedentary or family-oriented versus not, smoking versus not, you know, outdoors people versus sit at home and watch TV, kind of things like that. The more you can add and the more that it's deterministically linked, the more, I call it a true holistic viewpoint that you can provide. And remember, these guys are using this for marketing purposes, they're not using it to get FDA approval, so they're looking for those little pearls that are going to help differentiate their drug in the mind of both the patient and the healthcare provider based upon clinical specificity, but the more information you can give them, so that when that patient walks in and she's got blonde hair, blue eyes, and three freckles on her face, you go, that's the drug I need to use.

Amie: And that probably actually happens. So, when you talk about deterministically matching patients, what you're saying essentially is you're able to understand that the same individual in one database is the same individual in the other database by looking at things like social, like name, etc., which again, you said you never see, but the folks in, let's say a clinical informatics setting would see and be able to match, and that eliminates double-counting, for example, because you know you're looking at the same person and it ultimately provides a complete view of that patient.

Shelli: Right, because if you're not doing a deterministic link, then you're using a different algorithm or projective technique to do that, the potential to double-count does exist.

Amie: So, Shelli, you talked a bit earlier about some of the information that gets extracted from natural language processing, you talked about severity scores and things of that nature. As valuable as that information is, we recognize that it's somewhat novel, especially for researchers or business people that haven't used it. So, what in your experience do you see as potential challenges and how do you overcome those types of challenges that are associated with natural language processing?

Shelli: So, let me take a little step back, I was just at the most recent PMSA meeting and this whole topic of automated learning was huge. There were several presentations on it. And there's a lot of progress that's being made in this total area. But what I found really interesting is that these presentations were focused on automated learning on structured data. With natural language processing or NLP, the focus is on the unstructured data found in the physician notes. So, are there challenges with NLP? Of course there are, but that's why it's so difficult to master. Traditional NLP was not contextual in nature. By that, I mean it looked for specific words, it

didn't look for context, there was no interpretation, the word was either there or it wasn't there. But, in my perspective, as is most things, you get what you pay for. So, as technology increases, capabilities keep increasing, the ability for NLP to provide contextual information, to discern if the text is a positive or a negative, has huge implications on the interpretability of this data.

Let me give you an example. In the old days, if NLP recognized the word "severe," that's what you saw, severe, but the actual context was not included. So, how can you interpret it? I mean, you really can't. Is it saying a specific symptom is severe? Maybe it's less severe or more severe or not severe. The new technology and the new developments that are coming down the pathway and in actual reality, available today, is always being challenged and is always becoming more sophisticated. As is the same with all of the different EMR systems, they become more sophisticated, entering data becomes easier, the accuracy of data becomes more reliable and more complete. It's the same way with the NLP. So, with the challenges with NLP, it provides impetus for further improvement.

From my perspective, clients make decisions based upon qualitative insight on a regular basis. I have had clients that are in the back of a focus group of 10 physicians run out of that room and make a decision based upon what they heard from one or two physicians out of 10. Come on now, you know? So, most of the time, this qualitative insight is based on small numbers of responses. With NLP, these derived insights come from usually tens of thousands of records, which is far more reliable than 10 people in the room. The NLP-derived insights help bring that softer, more qualitative perspective to the patient record. The advances in sophistication, the depth of interpretation to the quantitative analytics just continues to grow. NLP helps us uncover test results that may have only been entered by the healthcare professional and not in the structured form. MMSE is one of those as an example. It's not a lab test, you're not going to get a hard-structured lab test back from it, it's a test that's administered and they may write down what they feel that that score was on that test. Do they all write it down? No, but those that do, then we have it. And while I believe the insight provided by NLP is more accurate than those provided by different recall techniques, the NLP also provides insight that could be utilized as clients further their own either qualitative or quantitative efforts with primary research. So, I'm a big believer in the NLP and I think, yes, there may be challenges, but there's no more challenges with NLP than we find in what I find to be our traditional qualitative techniques.

Amie: Right. And you mentioned the mini mental state exam, the MMSE, which is used in the Alzheimer's population, so I just wanted to underscore that

in natural language processing, not only do we see symptoms, but we also see scales and scores, so there is quantitative information as well as...

Shelli: Oh, absolutely. Yeah, I mean, if you think about the things that a physician puts into their notes, I mean, think about when your at the doctor and they're typing things in, most of that's not structured. It may be structured as to this is what I'm prescribing and now, finally, when they hit a prescription, they can actually get dosing, so our dosing information is much better, it doesn't have to be put in by hand anymore, but there's all that other soft information that goes in there. That brings a different perspective to what we're seeing in all of the quantitative part for big data.

Amie: So, that's great. Can you give me a specific example of a project where you've worked that had an impact on the brand team specifically?

Shelli: With NLP?

Amie: Within NLP, yes.

Shelli: Yeah, absolutely and this is a simple one. The client was interested in gaining insights into sleep disorders among Alzheimer's patients. I used NLP to pull mentions of sleep disorders among Alzheimer's patients and not just sleep disorders, but anything that had to do with sleep concerns or sleep issues that were present in the note. So, what do we find using NLP? Well, most physicians do not use the term sleep disorders in their notes, which our client had planned most of their messaging around using the term sleep disorders. So, what did they say? They wrote down things like trouble falling asleep, trouble staying asleep, restless sleep. Those were the kinds of terms that were used most frequently by the healthcare professionals as they recorded these symptoms. So, the brand manager was really excited. His exact words were, "This won't change how we go to market, but it'll totally change how we speak about the market." I think that's a great example.

Amie: It's pretty impactful.

Shelli: Yes, absolutely.

Amie: Well, Shelli, it's been a real pleasure having you on the podcast today. I really appreciate you taking the time and walking us through a more commercially-focused aspect of data.

Shelli: Great, thanks, Amie, much-appreciated.