Novel Cohorts
Podcast Series

Dan Housman, Chief Technology Officer of Graticule, and Joe Dustin from Dauntless eClinical Strategies discuss how EHR-EDC integration, AI, and interoperability standards are streamlining clinical trial operations

Dan Housman

Hello. This is Dan Housman, and I am here with another episode of the Graticule Novel Cohorts Podcast. I have with me here Joe Dustin, he's been a great industry expert in clinical trials, clinical operations, real world data throughout the years. I originally met him at a conference where he was working at Medidata at the time, and he was wearing about 12 different wearables as he presented, and you could watch his blood pressure go up and down. He's just a great expert and innovator in this space, and has been around the block from CROs to pharma companies to EDC companies, and I think he's now on his own. I'll let him introduce himself and talk a little bit about his background and how it shaped his thoughts around how EHRs come into play for clinical research.

Joe Dustin

That's great. Thanks, Dan. Great to see you again. It's been, it's been a while. Thanks for the intro. My name is Joe Dustin. I've been in this clinical research industry for about 20 years. I've always straddled the line between clinical operations and sort of the digital future, if you will, whether I was inside pharma or at a vendor company. I've always focused on using technology to make trials faster, more efficient and more human centered. I've spent a number of years at various companies in various roles. I've been in project management and operations, delivery, sales, product development, consulting, strategy, management. So I've held a number of different roles. I guess the highlights are, I spent 12 years at Medidata where you and I met, and a number of key roles there. Then I spent some time at Bristol Myers Squibb, as the head of clinical innovations. So that was during COVID. So it was a weird time. We tried to break a lot of things. But, you know, things were already broken because of COVID. So we had a very interesting time there. While I was there, one of the, I guess the two main things that I got done with my team was figuring out e source and the whole EHR to EDC thing in the very beginning of that movement, and then decentralized trials. From there, I went to Medidata for a few years, worked on eCOA, and then to BMS. We worked on digital biomarkers, and had even more wearables strapped to my body on various remote patient monitoring signals for real time detection and biomarker development. Today, I am an independent consultant at the moment, at least in 2025 and my firm is called Dauntless eClinical Strategies, and I just help life science companies, both both sponsors and vendors, navigate sort of digital transformation, especially around AI digital endpoints, eSource and clinical and eClinic technology that that technology stack that we've had for many years, it's changing. It's evolving. It kind of goes through, we're going through another one of those every 10 year cycles right now, I've seen three of these major transformations. One was on premise paper to electronic, on premise to the cloud, and now whatever AI is going to transform is substantial enough where it is a major, major paradigm shift in how we're operating in clinical research, how we operate clinical trials, how we find patients, how we move data, how we do everything. And so I'm pretty excited to talk to you about all this today

Dan Housman

Awesome. Let's dig right in. What do you think the pressing problems are? I guess we'll center ourselves in clinical trials, where electronic health records can help? Because, you know, we at Graticule and the RWE universe are swinging into the clinical ops world, but maybe we're all swinging together into the same shared space. But where do you think the EHR could help, and what's the problem where we could even solve it?

Joe Dustin

Well, there's a number of use cases if you can tap into EHR systems in the US and even globally, there's a number of things you can do once you're there. The EHR to EDC thing is defined as moving data from an electronic health record system in the healthcare world into a clinical trial data capture system in the research which is not the same, the data doesn't always map directly. The standards are not the same, and I think in most cases, for EHR systems, the data is not structured. So there is structured data, like labs and different types of demographics, medical history and whatnot, but a lot of that stuff is in unstructured notes, as we would as we would see normally. And so there's a few use cases we could use this data for. Number one is taking hold right now, is tapping into EHRs, where clinical research is being conducted, to avoid what is in Data Management world, to avoid what we call the swivel chair data entry problem at clinical research sites that are conducting trials and using the system that they want to use, which in hospitals and academic centers and whatnot, is typically in EHR, where all their data stored, and they don't want to have to go, copied here and then copied to another system, like an EDC. And when I first click a button, and it just moves over and grabs what I need, and then my workflow becomes a confirmed, confirmed, confirmed, confirmed, as opposed to typing it, and usually, probably will interns and other staff I need to bring on to do it quickly. So that's one use case. I think we'll talk a lot about that today. The second use case, I think, is for EHR, the matching of inclusion criteria and various types of demographics and medical history and medications and whatnot, for patient matching from an EHR as one of the primary sources for clinical trial recruitment. There's a few different ways to do this. There's the top of the funnel way and the bottom of the funnel way. Top of the funnel is using the basic data from an EHR that's exposed and matched to a protocols, inclusion criteria to then return a list of patients that you need to then go call or email or drive ads to or somehow get their attention, and then you have to do further pre screening on them and decide whether or not they actually can enroll, and they decide whether they even want to enroll at all. So there's the top of the funnel matching that can be done with EHRs once you get into the data and once you get into the systems. The bottom of the funnel option is sort of EHR record retrieval, where in the complex oncology studies, you can grab the data, which you will need, I think, for qualification and screening and sort of inclusion criteria checking, because they're very complicated right , oncology studies, you have to check a lot of information. You need the full records to sort of do that even at the top of the funnel, right? But doing that usually costs a little bit more money. It takes a bit more time. It is a complicated process. But there are, there are technologies out there that are being developed that are getting at this data in a much faster way. I've seen even in a personal situation, it takes almost a month to get some of these records for these reasons. It shouldn't take that long, especially if you're in a clinical trial scenario where you may not have a month, right? And the bottom of the funnel scenario is once this patient has been pre screened through maybe one of these other matching protocols that you then retrieve the records for everything else that you need for, just to get the records into the study for integrated

evidence, not necessarily a RWD data poll, but just the records you have that you need, especially in an academic or hospital research setting, you need those records to be transferred to the clinical trial. So that's three options for EHR. So we have the EHR to EEC, we have patient matching, we have record retrieval. There are other use cases, I think, in a real world data sense, where you're doing sort of more, not registries, but more of these, searching for cohorts of patients out there in the real world through EHRs as a way to just follow the long term. This is when we get into the topic of for some time called tokenization, which, when I was in pharma, we called it lifelong follow up. And you're then grabbing EHR data through tokenization, which is in the real world, just a blinded way to grab your data from the EHR after you've completed your clinical trial, and we can just keep track of how you're doing for as long as we need, without capturing any of your identifiable information. And so that's another way that getting into this more the EHR data across the world is being more aggregated now and put to use in those cases for not only tracking patients long term to see how a drug is doing in the market, but also maybe to help define and design trials in the future.

Dan Housman

Well, let's go back a step to this EHR to EDC, EHR sourcing world. You know, what's the benefit to the sponsors? What's the benefit to the sites? And I'll probably add, this question seems pretty simple. Why isn't it done? It's like it's been 30 years. You know, we have had things like Epic since 2008. It's 2025 . What's taking so long? Yeah. I'm just curious. Are the benefits not outweighing the cost to change? Are there certain barriers you think that are breaking down?

Joe Dustin

Yeah, there are a few reasons. Part of it is the perceived barriers to change are too high for people to care as much, and also the incentive structures are different. Right? In clinical trials, we have a very unique scenario, which is that the B to B scenario is more common than not, but in our use case, it's a bit different. The people who buy the technology, the people who fund the technology, and the people who use the technology could be three different people, versus the direct to consumer, where the person that's buying it's going to get all the value out of it. And that's an easy exchange, right? But in our world, you might have a sponsor paying for it, or CRO implementing it, a site using it, and their incentives are completely different, so the site wants to use it to make their life better. The pharma companies like, Well, I would love to make your life better, but really, I don't care, because I'm just going to pay you to do whatever, and I want you to do it my way, because we're trying to maintain the integrity of our data, and we need to prioritize what we need. You've probably heard of the topic, like the sponsor of choice programs right to sort of work better with sites. And of course every pharma company would tell you, especially even CROs would say we need to have good relationships with our sites. That's what we have.

Dan Housman

So you think a lot of the challenges are just this split of who buys it, who uses it, and it's just also some of these incentives, we think would naturally be there, but there's, there's, there's a challenge. The incentives aren't 100% aligned. Are they aligning better now?

Joe Dustin

I think they are aligning a little bit better, right? So we have sites in a sponsored choice scenario, where sites are in a competitive enrollment scenario. That is one of the incentives that matters, because a pharma company is going to want to make it easier to work with them, for the sites that are just going to give their patients to another sponsors trial that's easier to work with, so if you become the one easy to work with, I can enroll fast. I can click a button and get all my data in. I can use a number of more modernized tools to make my workflow more efficient. Or I can even use my own systems that I already have in your study. I'm going to prefer to work better the way it works better for me as a site, even if you're trying to give me stuff, sites are still getting on average, like 20 plus systems, not for study, but overall, across all their sponsors. It's just way too many systems to log into, and it's getting ridiculous. And so we need to reduce those. The EHR barrier has always been data standards, data accessibility and standard is not data standards, but standardization of how we acquire data. In pharma, they cannot, and I dealt with this little bit when I was at Bristol Myers. They have a real problem with optionality and flexibility, similarly to how we made, you know, talked about decentralized trials as the key to that was optionality for patients. Optionality for pharma sometimes isn't looked upon as good, because they want one SOP, one process for everyone, and then they just do that, and they abide by it, and there's any deviations, they record it, they log it, they handle it. But there's always one way to do it. If you have the sites tapping into their EHRs and acquiring data that way, you might have how many different processes for how many different sites, and then, is it a custom integration for that site? We have to do that every study, or just once. It's just these things like, well, we don't really know the answers to that, and we don't have time to deal with it. I'm not putting it on my study timeline to figure it out, so we're not doing anything. It's usually down to timelines of a team that's running a study. No one's figured it out in a standard way, centrally, and they just don't want to deal with it in their study. That's one barrier, right? It's people, it's time, it's process, it's priorities. But no one would mostly, no one would disagree that it's a good idea, right? One would say in some cases, the data quality would improve because it'd be less queries, less deviations, less issues, because it's coming straight from the source. Others would say, I don't trust the data in the EHR at all because garbage in, garbage out. But I would also say that it depends what you're using it for. If you're taking retrospective data, we all know any EHR is like, a lot of the data in there is meant for billing purposes, which means there are things in there that aren't necessarily as accurate as you would want, because they put the codes in and the right things in there to get the insurance to cover it, and necessarily because you have that condition as coded right. And so there's always up and down when it comes to data quality and data integrity, but, but that's for more like screening and matching and things like that, for prospective data entry in the EHR for the trial that's usually legit. Whatever's in a structured data field goes in and maps to an EDC, whatever is an unstructured note can be grabbed now by an AI or some kind of NLP type tool to go through the notes, grab what's necessary, and start to map that and classify it for where it would go in the trial. And so the barrier has been that technology has not existed, not been easy to use, has been expensive and not scalable or widespread, but now it's changing. I'll give you a couple of reasons why the barriers are coming down. Right? Number one, FHIR HL7 is the standard in healthcare interoperability, portability of data under HIPAA, right? That you need to be able to get your data out the HL7 messages, how lab data gets around to different hospital buildings, to how you can get into your lab portal, etc. And also is how you can get your healthcare data anywhere. Research doesn't talk HL7, although there are some pharma companies that are starting to ingest HL seven data directly, and skipping over CDISC, if you will, CDISC, ODM anyway, to go straight into their ecosystem, because they want to start speaking the language of sites more natively. Two things happened. Number one, FHIR HL7 reached a milestone version update a couple years ago. FHIR HL7

V 4.01 I think it was that was mandated by the US government, that every hospital, or anyone getting federal funds or something under the one of these, one of these acts out there just escaping me at the moment, you will be fine if you don't have this up to date. So most hospitals, especially the big ones, updated everything. And do you remember, like, two years ago around Christmas time, when all of a sudden you could Google search some doctors and like ones in California, all of a sudden you could, you could schedule a doctor's appointment on a Google search, just like you did a restaurant reservation, like in the Google results page, you could schedule a doctor's appointment. It's because of that update, data was made more open and interoperable, and that's how the real world data companies, the aggregators, all of a sudden, their business started going up because they were able to get access to much more data. As a result, interoperability within EHR and EDC just got easier, not because EDC could talk HL7, but because more and more of these data are becoming more structured out in this clinical trial, aware, if you will. The third thing is a project at HL7 called Project Vulcan. Project Vulcan is meant to make most of HL7 clinical trial native, so that you could have the ID of a protocol, ID a site, ID the inclusion criteria, schedule of events, the table of scheduled assessments all natively understood in an EHR. And so that just makes a clinical research trial in an EHR that much easier. And that's still a couple years away from getting there, but that has made strides. So the access to data has become easier.

Dan Housman

So it sounds like there's least. I love your picture of these different waves, this, this new wave, you characterize it to be this AI wave, but it's also backed by this interoperability wave that popped up a couple years ago. So we're in, you know, we're in a combination of events, I wouldn't say, necessarily, The Perfect Storm, you know, because we're in a moment that moves slow, but Right, right? These two things are both necessary, right?

Joe Dustin

Yeah, you need that. If I make a comparison to the first wave of clinical research technology innovation in the early 2000s right? You had electronic data capture as one of the major moves that transformed the way clinical operation works. It was and wasn't right. It was because it put things online and allowed clinical sites to get their data online faster, which got sponsors to get the data faster, which was the primary reason for EDC, what was the main value point we sold to Pharma. It reduces your time to the database, right? That was the main thing, got it and it, but it floated on top of other IT transformations. It didn't just do EDC because it was a good idea. They did it because they're already going paper to electronic in IT and moving to the cloud from these on premise, software service started becoming a thing. Budgets were being redone. If that wasn't happening, this never would have been adopted right as fast as it was, and it took many years to get there. But in the early 2000s I remember, was it 2006 when I was Medidata Amgen was one of their early clients. I remember in 2006 they put out a press release saying that 90% of their studies will be in EDC in one year. When Amgen does something, they put their entire company behind it, and like nothing else happens until that gets done like a major in ClinOps, right? They've done this a few times with protocol design and other types of technology initiatives, but they did it. They were successful in it, and the entire company went EDC in one year. They got it on every single trial and moved there. And that was early for the industry, if you would. And they were doing it the new way, right. In those early days, they were still shooting laptops on premise the old way. And, you know, a couple crazy guys

from New York Medidata said, we're gonna use the open internet, because that's where things are going. We're gonna do it in a browser, we're not gonna install anything. And that was like, You guys are crazy quick. They were already ahead of the game when it became normal, right? And that's the kind of thing we're starting to see now with this next floating on top of an infrastructure change that's moving more toward an AI driven build, design, extract, you know, use case of the data analytics also. And the EHR to EDC thing is happening. And there are other things that are riding the wave of that as well, to get into the data and make collaborations and data management more efficient, reduced cycle times, all that kind of stuff.

Dan Housman

So it sounds like whatever Amgen did worked great with EDC. I'm curious how you think pharma should go about adopting these capabilities. Should they take an Amgen approach? I think it is slightly different, which makes it challenging that there's other people involved, right? So if Amgen is trying to adopt EDC, it's like, Hey, here's our EDC system. Now it's, how are we going to interface with all these decision makers who control this really powerful asset, but also a very restricted asset called the EHR? And can you do this in a year, is it? Should it be done incrementally with pilots? How do you feel about what the pilots have done in the past versus what they should be doing going forward? What do you think of the adoption of you if you were now back in a pharma company like you once were, what would you recommend they do?

Joe Dustin

Great question. So I'll use it in my consulting practice working with some pharma companies that are doing just this right. And so there are active things happening, and most of the major pharma companies now have an initiative for what you would call EHR to EDC, which is tied into multiple initiatives at a digital strategy level, right, dealing with variable data, dealing with recruitment, dealing with just enhanced data management. Their data management departments are migrating, and they have been for some years, from a clinical data management group to a clinical data science group. There have been two different groups at the same time, but their roles are changing. Right? The role of data management has completely changed. The Society for Clinical Data Management put out a paper, and this is 2016 on how everything's going to change over the next 10 years, and we're pretty much there at this point. So they were, they were spot on, because the robots are doing the work now, and for good reasons, we're going to do things way faster, but now that people are doing different jobs, right, which is how part of how the adoption changes, because in the EDC world, if I make a comparison, we adopted technology, but the process, honestly, the process was horrible. It didn't change. They were still collecting data on paper at the sites. They were still using binders, and they were still transcribing stuff into another system that pharma gave them. They didn't change the way they did anything at the sites. They just, instead of sending the paper out for someone else to deal with it, they made the sites do it. They were right, and the adoption of that was faster because we didn't really disrupt how the sites worked. But pharma got a benefit of getting things faster by paying the sites to do a little bit extra work to get the data up online so they could do the review and queries and SDV and all that kind of stuff. Right now in this paradigm process, we can't, because now sites are starting to adopt their own technology, which is meaning they're digitizing their own operations for their own benefit, hospitals, academic centers large sites already having an EHR because they treat patients, and they see patients, and they have a cancer center, and they do

whatever they do a research hospital. The sites that don't have an EHR are also digitizing their own operations. So your smaller, independent sites, your site networks, those that are in different parts of the world, in different parts of the country, even those in rural areas, aren't tough to get to. And so the adoption of that is increasing, so that you have more sites coming online and ditching their paper process and changing their process, which now allows this new connection to happen. So we go back again. So, this is happening on the side, right from what comes to pilots and how to adopt it, when a pharma company wants to start doing this. When I was at BMS, we piloted, we did 3 to 5, or 4 pilots, in a sense of within a year, I would say maybe a little less than a year, we piloted three different EHR to EDC connections, I'd say one was a spreadsheet only in a mapping exercise with data management, and two of them were in live studies. Were a couple of sites at production studies, right? And then it just got bigger from there. They ended up doing about 10 to 13 studies or so in that model, as far as I know, still going on, but they saw massive cycle time reductions in data management and a cost decrease there. It was like a 44% cycle time reduction by getting data faster, by having it higher quality and accurate the first time, and they were just doing source data review, not source data verification, right? So a lot of that stuff went down, but it took a number of pilots and a number of realizations in what tools are doing, what and why, and what sites were worth it. So the first time we did it, a lot of these companies were claiming here, you'll get, like, 60, 70% of your data from an EHR into your eCRF, cool. The first time around, we only got 18% so not at all. So it was a high expectation, low delivery, not because the company wasn't any good. It's just that the data was not what we needed. And also a big, a lot of these EHR to EDC initiatives, the big percentage of data is and the workload reduction is based on lab data. That's in the EHR right? Especially in oncology studies. Did you know It takes about three hours per patient that a clinical coordinator has to deal with just to get that patient's data organized? Most of that is because of the lab data and all the labs that they take. What they take, need to grab and how they're transcribing it into a portal from a CRO which then is manually organized and harmonized and delivered to the sponsor. In this case, they spit the button any data goes. But in some cases, the pharma company doesn't want to put their lab data into EDC. It's got to go somewhere else. And oh, that is an EHR to EDC feed can't put the lab daily in on the other plate inside the EDC,

Dan Housman

There's nowhere to put out. Like the EDC doesn't have it, or its

Joe Dustin

EDC systems have lab modules, at least the two big ones do, but a lot of them don't support HL7 right? So you can't get the message to come over. It's like they're used to, like batch loads from Central labs. It's not the same thing, in which case these EHR to EDC feeds in the early days, this is like 2021, we were doing this. It was to one where it might go to a data lake. It might go to some other staging environment where it's then, you know, conformed into a newer, standardized format, right? And so the adoption of this was through many pilots. We realized some of the data we're not getting as much as we thought so, and that's for certain kinds of studies that maybe it's not worth the effort into sites aren't already having an EHR there are other companies we use where we did get about 60% of the data, and the reason we got the 60% of the data was because they went after the unstructured data. Also the early EHR, EDC companies and many of the ones today still are only grabbing structured data from a FHIR HL7 feed. And it's not enough. It's not enough. We need

unstructured data to make any difference in the real value to what Pharma would measure as that worked, and that was one of our times, right? So now, the AI and the ability to grab that data and parse it through and then categorize it and map it and go, it's possible, and it's happening. And there, there are companies that are doing this now, and they're only getting better at it.

Dan Housman

Yeah, I guess I'd add on this one, Joe, this is from our own experience. You know, the reason why unstructured data is hard is because it has PHI in it, and everybody's afraid of the PHI. It's like toxic waste. Don't, don't touch this thing, because if we pass along PHI to a pharma company, now, everybody's got to have a BAA. And it gets complicated. But one of the big benefits of AI we've seen in last few years, and we've employed it in the work we're doing, is we can get HIPAA certified great sub components to de-identify the unstructured data before we process it with all the other AI, meaning we can move it, which is actually a huge piece, because we can't move it, we have to move All the AI to where the data is, and that gets complicated, right? Because moving data is hard enough, but moving AI into data creates its own issue. And so, you know, we're enjoying the fact that now we have this new set of components where we can just bring in all the unstructured data, put it into our cloud and operate against it with the pharma company, right? Instead of having to ask the site to go operate against it independently,

Joe Dustin

It brings up an interesting question about the data harmonization and who should do it. Example, the days of early days of eCOA, where I was doing patient report outcomes on pilots, and then eventually to smartphones and the web, those disparate point solution companies that only had eCOA and they would always get a data management spec from the sponsor and say, and you just send me data in this format, usually it's an STDM format, and drop it on a file drop like a FTP site, and then I'll take it in. I don't think we can do that with sites like that. It is not something that can scale to every different data source, right? Sponsors that think are wrestling with the idea with, if we are going to go from an eSource system or an EHR system or something, we should be able to take the data and whatever comes in and, you know, spin it around in a in a centrifuge, and pop it out in the standard format that we need. But do we need to ingest everything to do that? Or do we still only ingest what we need? you need context, it's like bringing the AI to the system. Can you do that all there and then bring it in? Or do you have to grab it all first and then spin it any grabbing things that you should be right, and it was a lot of data in the EHR. You mentioned PHI, you want to avoid that as much as possible. And then in eSource systems that are also like site CTMS, type systems that also have data capture, direct data capture capabilities. Those systems don't speak FHIR, they don't, they're not in the EHR, but they also have the ability to be extracted from and pull data into sponsor infrastructures. They could speak CDISC ODM at the minimum. But should they? Should they also be following through the same pipe as the EHRs, in a way, when I think about adoption and standards and bringing the AI in what direction I almost see again, because the infrastructure is changing before our eyes, and other initiatives are floating along. The same transformation that's happening. You're going to have the EHR to EDC thing happen. The AI is probably going to make that faster, and it's probably going to be the sponsor side that grabs it and transforms it. Then you're going to have these eSource systems that are not EHR, that are going to be all also flowing through the same pipes, because pharma is going to take a year to validate some connection that they want everybody to just

use. Some pharma companies are going to say, no, we'll just have five different pipes from we just want whatever the sites are using, and we want to be able to acquire that in a flexible way, right? We want to make it easier. There's going to be both. Both of those are going to happen. Both those methods are going to be adopted. And depending on the company and their initiative and their digital strategy and the resources they have, they'll go whichever way is the path to least resistance and higher quality, right quality will always win. And then it's time and cost. There's an initiative inside the DTRA, the Decentralized Trials Research Alliance that I'm running as the workstream lead, with other colleagues of mine from pharma and tech, called BYOT, Bring Your Own Technology, and that is putting a framework together, releasing a playbook in July 2025 releasing a playbook that will basically be the how you get a yes from pharma when a site asks, Can I use my own system in your study and not the system you're telling me to use? And that would be eConsent, eSource, I don't want to log into your EDC ever again. Maybe, maybe the PI signs stuff there. But like, I want to use my source system, and I want to send you data, and that is probably gonna maybe jump and, you know, piggyback on, on the pipes that the EHR data is going through as well. But what's happening now is these companies are deciding, like, should I make the eSource systems, FHIR HL7, seven compatible. Or should I just be able to create another pipe and ingest from the sites that don't have an EHR but the EHR initiatives are driving the transformation. And should we be thinking wider? And it's all related in some way, right?

Dan Housman

Well,I certainly see this, you know, this dual hat problem on the pharma side, which is I need to be able to do what the sites say they're going to do, because I don't control the sites. But also, I need my system to be reliable in exactly the way I want it, because I'm inflexible, and it's how I have to run a study, so I need to tell them to do it my way. And I think the BYOT is an interesting framework to put it in, which turns things into APIs and standards, and as long as you meet my requirements, it's fine, right? You've created the capability requirement. And I'm gonna ask you a question here, which is, I feel like in a lot of what we're talking about here, we start invoking the need for some middlemen. And like I would, I will call them honest brokers, and the honest broker might be the technology provider that's providing the eSource or the framework, or it might be because the site itself, this ain't their main business, and a tool like Epic is never going to do these things for them out of the box, right? And so for them to take their IT team and build out a whole infrastructure, which I've seen groups do, right? It is a big undertaking, but it's not going to scale. Well, it's going to work for the I'll say, like the MD Anderson world, the MGBs of the world, they can go and invent because they have the budget from whatever the massive research enterprise they have. Everyone else is like, I can't do that, right? I'm going to need middlemen to go do all this work for us. So I'm just curious what you think about the, you know, is it an honest broker? Is, you know, even back to this question you had about you can't send all the data, because it could be a problem to send all the data that you don't need, right? The EHR could just go throw out. Here's the full record of this patient. And I think a lot of life science companies, you say, companies who say, I don't want that, right? Yeah, I want to be exposed to the subsets I need. And so who's going to be in between? Because we put on the site, they're like, I have no way to subset this thing for what you need. That's not what my team does to create a whole IT department for each study that you know, basically make a little machine to slice the data down by hand for your custom study. So I'm just wondering how you think it's gonna play out.

Joe Dustin

The honest broker idea is interesting. Let's dissect this a little bit so that you have these sites, hospitals, small sites, big sites, whatnot that have data. The data is being sent across electronically to various locations where the connection is available. There are real world data companies that are extracting data, and they're either paying sites for their data, that sites basically selling their data to, like, you know, real world data companies that are aggregating it and selling it to pharma and then, a result of these big hospitals are getting, like, either appliances installed in their server rooms, or they're getting reports and stuff back as a result of something that they get value out of them. So that's how the real world data companies are getting their data being extracted from the sites, right? And so in another honest program, but they're a third party between the site data or the use case of EHR for research. Are you suggesting that maybe there's a need for, like, a health information exchange type thing for research data pulling out?

Dan Housman

I think, I think that's what comes to mind when you say these things, that the process of appropriately managing the data is one part CRO, one part technology, one part RWD , yeah, right, that that's not, it's not like, you know, any one Hat would work if it's just a pass through, which actually is like an HIE doesn't do a lot of manipulation of the data. It sends through great, right? They might make it available, yeah, so as you just get the data in a better format, right? Yeah, they'll clean it up a little bit to make sure it's nice. Maybe it's nice looking, but they're not. They're not going to go in and say, I'm only going to send you the six fields you need. Bob, I that's your right and so, but here, yeah, it's more complex, right?

Joe Dustin

This honest broker thing, I think, does depend on where you are in the world. I think that if you were in a socialized country with socialized medicine, single payer, it might be easier, because that data is technically coming from one place, maybe a couple places, that's government controlled, state controlled, or in the US. I just don't think it'll ever work anytime soon, because of this and even across Europe, GDPR has very strict regulations on what to store, but they still have countries that are socialized or they can get at this data. You know, everyone has a national health ID. Everybody has the ability to get their data easily around where necessary, and people can find it and get at it with access much easier. Right in the US is just so fragmented that even the the thought of anybody trying to aggregate all of your data for research purposes freaks people out from a privacy perspective, because

Dan Housman

I was thinking more of an honest broker in the workflow for the patients already consented to a study that neither the pharma nor the health systems in a good position to perform the filtering role. They both could conform to the transfer and receipt role. But there's somewhere in between where they're like, you know, I guess we can have our people manually do it. I'm just thinking about study with 50 sites, and you say, hey, each site, I want you to only give me this subset of your data. And

now they're like, well, now I'm building like a robot at each site, which is going to look exactly the same, but they have to do it from scratch.

Joe Dustin

Well, to your point earlier, about large pharma companies having the big budgets and large sites with big budgets to be able to do some of the stuff for the smaller folks, that would be quite a service or a benefit, I think, to the smaller folks that are not going to pioneer that themselves, right? So if you're talking about an honest broker, sort of data marked for consented patients in a trial that can be spun up on a per trial or a per company basis, I feel like that is an unmet need that no one really has yet, so that's something that maybe I should work on, that

Dan Housman

Well, we'll see. I mean, I'm interested to have it all play out, because I think I see a variety of models coming out, none of which have taken off like a rocket, right? So when I think of the steam engine, if you ever read this book, there was a book about, like, the history of the steam engine, and the first thing they did with the steam engine is they made a tool to mine coal faster. And what it did was, because coal powered the steam engine, it had a net positive gain, and they could just mine coal tons faster and use the steam engine like, hey, we use this for all sorts of stuff. I do all sorts of stuff, right? Because it had this net positive. So if you know, and I think we were still lacking the invention with the EHR data for clinical studies, that's like a steam engine, right? It's got enough net benefit that people are just going to keep expanding it on all sides, sticking a train,

Joe Dustin

There's more value you can extract from that data in a number of ways, like we were talking about, mostly moving data across, dealing with the management of that data, the harmonization to make sure that data is clean and then used in the right places, to make the site experience better, to reduce cycle times, and data management to speed up the acquisition of that data, which then speeds up the management and the use of that data in some kind of analysis. Like, I could see a world where this is flowing like, the need for an interim lock is almost not necessary because it's just happening in real time, like, all the time, right? I could also see the use of AI replacing a lot of these edit checks that people would be typically putting in EDC because it's because you needed to check the data to be manually transcribed. But the stuff is coming automatically from the source, and the source is already dealing with that kind of stuff. You don't need the same kind of management and date of data that we have in the past. The entire role is going to change, which will reduce cycle times, reduce cost and increase the speed, not only the database law, but just overall trial operations of getting this stuff done. I also think that when you talk about the use case of it, just to flip back to like patient matching again. I've seen hundreds of companies claiming to do some kind of patient matching service at this point, whether it's technology driven or service driven, so many of them are now, you know, magically AI powered all of a sudden. But really what it is is we can match patients all day, whether the site takes a tool on their own, uploads a protocol to some AI tool and then mines their local database of patients to find matches, and then sends out like an AI voice bot. I've seen some of these now that are coming out, that will call patients and interact with them over chats or voice and you can ask questions about the trial. You can ask questions for pre screening and get

employment super fast. Pretty cool, right? Imagine hitting that like a list of like, 300 people that you match and just hit the bot. And then, you know, make some of your first calls, but especially the second or third time you try to call these people, they don't pick up, send the bots out and see if you get a different response. But it's still, there is no guarantee that these tools improve enrollments, right? That's still a behavior. That's still a behavioral problem, right? That is unless your family or your doctor or some promise of free treatment or cheaper treatment, or something is made to you and you think it's a benefit to your life, you're not going to do it. And so that's where educating about research and making it more aware about what it is and why it matters, why it matters to you, is a very personal decision. And so with the increase of data availability and the sort of improvement of AI our tools to match patients only go so far, and sponsors really aren't buying it like these tools. They're like the sites where you can buy it. You can reduce your workload. That's great. I only care about if more people are enrolling as a result, and that's not necessarily the tech that's doing that. That could be the way your site operates, the way you deal with patients, the way you get a really good PI like, it just depends.

Dan Housman

I mean, it sounds like, because we're also somewhat skeptical, because we've worked on some projects relating to recruitment through AI to find the right list. And it does seem like the problems are not like you can't solve them with just the AI. A lot of this is, is the referral pattern correct for the patient when they see the right doctor? And is that the doctor that's supposed to talk to them about this particular kind of study to make it work? I guess I'd argue maybe a lot of the AI companies are just naive, right? Like, we just throw AI at this problem, but it's like, it's a social problem that you have to also throw the AI at. And, you know, people think, like, sorting hat's the problem, right? Gryffindor, oh. Gryffindor, we found a Gryffindor good. We're good. Houston, Gryffindors, we're fine, right? But the reality is, like you're not, you're not at Hogwarts, right? Yeah, you have, like, a lot of other options here, right? You could just not be a wizard at all, right? Or,

Joe Dustin

When I've talked to a lot of companies, they're doing this, they take a very boots on the ground approach and human touch is important like that. Connection is important to merely really get trust from patients, especially in more diverse communities and people of different socioeconomic statuses that aren't in the big cities, that don't really trust anything that to understand that this would be helpful for them. You just need a different approach in some cases. And the technology approach is not always the best. I think it's helpful tool to make the process go faster, but in some cases, this shouldn't be the first thing to do, just because it's cool,definitely helps,

Dan Housman

And maybe the place where you should fix it. And I was talking to someone at a big health system where they said, well, pharma always comes at us, to us and wants to fix the trials midstream. And this is the group that applies AI and has all the data. They're like, it's too late. However you wrote up the study, however you designed your recruitment strategy, everything you already did, the system behaves exactly the way you designed it, and isn't working. I cannot suddenly apply data and change your system. You know, you made a, you know, an inclusion criteria that was too restrictive, or

whatever it was. Maybe the right time to apply a lot of this stuff is just before the trial is starting to get contracted, or at least when it's being planned and like, apply these things so that you're ready at the get go, thinking through all the problems we just talked about. Right? Are we picking the right PI? Are we picking the right practices? Do we have them up, you know, up and running with the right lists and go,

Joe Dustin

Yeah, I have seen sponsors double their enrollment of the use of some of these EHR matching tools. I think it is unique to certain indications where chronic diseases and stuff like that, where it's more effective, for sure, and usually a high volume of patients that are in these studies, has an effect here that is actually positive. So there is promise for some of these tools. So keep watching. As things evolve. I think it's going to get more interesting, for sure, but the next two to five years should be pretty interesting.

Dan Housman

I know we've been on for a while, and I'm sure we could cover many, many, many more topics. Any closing thoughts, you know, maybe some vision. What do you think people are gonna be seeing in the next year or two? What would be big stuff on the horizon that you can see that other people may have not looked at yet?

Joe Dustin

All right? Two things, right? So there's going to be continued investment in, sort of the infrastructure around interoperability. It's going to be standards work, it's going to be technology, it'll be AI powered this and that, which is the infrastructure on interoperability, will be continued to to evolve and get more investment, both from pharma, both from the healthcare enterprise. And I would even wager to say from the governments in different countries that are going to have more centralized initiatives to make this better. We've already seen from the FDA, the new FDA, if you will, better, new priorities that I think are better than other people might question certain things. But one thing I will say is that the new heads of the FDA are are very they want to be very innovative, and they are making changes of things that we've been asking for for a decade, for them to be more open to and now they're starting to do this themselves, and they're going to use their things themselves, and you're going to start to see review times go down in theory. And I just want to be clear that the things the FDA are doing right now, regardless of who the boss is, is good. We should be looking at these things through a clear lens, and not through

Dan Housman

Do you think the FDA is going to prevent leadership positions here? Because I feel like they've been sort of very neutral in EHR, EDC, or EHR, into these things, they said, Hey, it's okay. Here's some guidance. Not, yeah, no. Like, if the FDA really wanted to make it honestly, if the FDA said, we mandate you have an interface to EHRs by 2028, it's done. I know, I know, and I'm not going that far.

But I'm saying, I think that pharma takes so much guidance from the FDA, part of the reluctance to move forward is the worry that the regulators won't accept doing it in new ways.

Joe Dustin

So here's another thing on that one, right that's going to drive a lot of this. So I put an article on LinkedIn right around the time of the DIA conference, mid June, that took an analytical look at the priorities, look at in your show notes later. But the one of the main things I saw this, it's like wasn't, wasn't explicitly defined in the paper that FDA, FDA published a paper in the JAMA Journal on their new priorities, right and but the thing that I caught away from that was the FDA is just being more open to these type of tools, which means, in general, the fear of being rejected by somebody who doesn't trust any of these things. They don't know about it, they're not using it themselves, it's just an overarching thing that Pharma has. I mean when they ask, and I'll get a clear answer now with them using these things and implementing these things themselves, and being asking to see more of these things that fear is going to go to that just inherently, and you're going to see more and more of these approvals happening as a result of, you know, it'd be really creative steps

Dan Housman

.Yeah, I'm like, nuts would be the FDA does run their own trials. If they mandated their own trials, they had to use these kinds of technologies. Because instead of being a laggard, they were a leader. Pharma would take note as well. Would everybody. Because I do think it's interesting. You said the reason why we have interoperability, health systems actually have not wanted interoperability. It's really good for patients. It's not good for health systems economics when they have to add new infrastructure to the interface that isn't paid for, and other things. So the only thing that seemed to have gotten it done is the regulators came and said, do it this way. And I do wonder in the back of my head that it'll take the regulators to push a little bit, to get a bunch of, like, squishy wheels to sort of smooth out a bit, because I think that it always does, right?

Joe Dustin

I mean, we've had the latest guidance for the ICH shall revise their E6(R3), which we're at now. So that's making changes that are going to cause people to move. I mean, when revision two came out years ago, we got risk based monitoring and all of that kind of happened. And when it was sort of mandated, people did things, and it was good for the industry. So look for more of that. So investment in infrastructure, I think we're going to see regulatory acceptance of more of this stuff. It's not gonna be a mandate, but you're gonna have more and more acceptance of this that will help make more decisions on the farmer side, more concrete, and make increased adoption. You're gonna, if anything in regulatory we won't go into this deeper, but you're gonna see some deregulation, which is also gonna help in what we're trying to do. I think you're going to see a lot more of these AI co-pilots, if you will, for sites and CRAs to help them do their jobs faster and more efficiently. They're going to be doing some of this stuff on their own without asking, right? Which is going to be a question, like, I'm putting my protocol in these tools. It's not my protocol, it's their protocol. So is that okay? And you know, who do you have to tell about that? That's going to happen, right? It's going to help the sites do things faster. It's going to help them find patients faster. It's going to help them reduce their workload. But what does that mean for, you know, privacy and whatnot.

Dan Housman

Co-pilots are kind of like these agentic models, where it's, yeah, sit next to this thing, and it just does the thing I normally would do, you know, it's maybe the next generation of RPA robotic processes or other AI processes. And it's like, yeah, you can do this, but I can do it too. I can impersonate you and go through all the same things you could do, right?

Joe Dustin

We're gonna move from data forward, some data collection, which is what we do today to sort of data orchestration, right? And this is, those are the final thoughts I'll leave you with, which means, as we're grabbing data from different sources across different sites, both in the US and around the world, because, like Europe, is starting to adopt interoperability standards that we are FHIR HL7 Japan is doing it, and so like the best standards that we are using for interoperability, the rest of the world is starting to pick up on it, and they're adding their comments as well. But that will become a global sort of thing. Now it's going to be orchestration of all these different sources coming in and normalizing and making sense of it to just speed up the entire research enterprise. So that is a massive change you're going to see in the next few years. See things just the beginning. And we've been talking about that since 2012, right? And back then I have myself on stage at DIA and point club saying, this is the next thing. Finally, they were finally making progress. And I just laughed at it when I was there, because we've come such a long way since then, in the way the standards have changed, but not in adoption. So I think now we're finally at like, a first real adoption spike, and it's only going to get better from there

Dan Housman

yeah, it's gonna be exciting to, like, replay this and see how wrong we are and where, you know, it's always these questions of, like, do we think far enough or near enough. And I'm sure something amazing is going to happen, and you know, medicine is going to get better. So thanks so much, Joe. I'm sure we'll all see you at scope and DIA and all the places that I end up bumping into you at, and maybe we'll talk again, you know, but this has been great, and yes,

Joe Dustin

you've been using here, and you know, the fall, I'll be on the road, and most likely see you and many of your listeners, so I look forward to it.

Dan Housman

All right. Great job. Catch you soon. Thanks so much. Bye. Bye.