Novel Cohorts
Podcast Series

Dan Housman, Chief Technology Officer of Graticule, and Melissa Haendel from the University of North Carolina discuss the challenges and potential solutions for rare disease research

Dan Housman (Graticule):

This is Dan Housman here from Graticule with the novel cohorts podcast. I have the pleasure of meeting today with Melissa Haendel from the University of North Carolina. We've worked on various projects relating to N3C (National Covid Cohort Collaborative) and there's some very exciting things going on around rare diseases that we wanted to talk about. Maybe first, it would be great for everyone to get an introduction to Melissa, and I'll let her describe a bit of the work she's done on national initiatives and other things relating to patient data for these kinds of areas.

Melissa Haendel (University of North Carolina):

Thanks so much, Dan, it's really wonderful to be here today with you. I have a little bit of a circuitous background in being having been trained originally as an early gene jockey, developmental biologist molecular genetics, and ended up working at a database where I was responsible for structuring phenotypic data from the zebrafish, and devised a method to compare phenotypic information from model organisms and compare that to humans. And thus was born a consortium called the Monarch Initiative, which has for the last 15 years or so, developed standards and technologies to leverage a wide variety of data sources in order to help improve the diagnostic efficacy of variant prioritization algorithms for improving rare disease diagnoses. Since then, I've also led several national initiatives, including the National COVID Cohort Collaborative and, more recently, the Center for linkage and acquisition of data for the All of Us, Research Program where we have collected and harmonized a wide variety of electronic health record data sources and link those data sources to other data sources, like claims data or mortality data, or geo coded environmental and social data, and through both of those kinds of initiatives, we can really start to paint a picture of rare disease longitudinal characterization, to try to better understand the journeys that rare disease patients go through, and start to really pick apart how we might have earlier interventions in their diagnostic odysseys. And so that's what I believe we are here to talk about today.

Dan Housman:

Great, why don't we talk a little bit about the problem? You know, what are we talking about? And I'll maybe start a bit from the pharma side, and I can ask your opinion as well, from from your perspective, because, you know, we work with a number of life sciences companies, and they come to us this challenge of, it's very hard to execute research on rare disease because the patient populations are hard to find, and often they're undiagnosed. And if they are

diagnosed, they may not yet have structured information the way you see it in a common disease. So it's easy enough to find someone with an ICD code for diabetes. You get into these rare diseases that have genetic components, there's often a lot of the key information for their product or their study relates to a certain genetic marker, or those patients don't get diagnosed because there's no treatment yet, because there's no point in putting a diagnosis in the medical record if there's no next step at the level of how to work and that end impact is a lot of inefficiency and a lot of missed opportunities to either get a study done effectively, or after a product's already in the market with an FDA approval, to have uptake, where patients benefit. And so we're, you know, we're, we're keenly interested in how to solve that. And I'd say the extra piece which makes this hard in rare disease is implementing anything that works is hard to scale, because you got little bits and pieces at each site. And there may only be 10 patients, 50 patients, 100 -300 patients, who will benefit from putting in new infrastructure. And so there's is the juice worth the squeeze problem? Half the time in any given place, there's no solution yet to all this. And you know, I could, I could probably go on and on within the real world, data world, it's also hard to do research, because you're looking at 400 patients out of a database of 100 million, and you're paying for a database of 100 million. So 100 million. It's even the business aspects of how to operate efficiently are hard. That's what we see a lot of, I don't know what you see on your side from the academic medical centers, government funded research, but I'd be interested to hear.

Melissa Haendel:

Yeah, yeah, those are, there's a lot of problems in there. And, you know, I think we the meeting in San Francisco, which, you know, the ARPA-H, new government agency focused on sort of high risk, high reward research, you know, you know, had a performers day for potential applicants of this program, which is focused on building a really robust data set for over 2,700 rare diseases. For each rare disease, collecting data about a number of different patients that have that rare disease, and using that to help people build diagnostic models. And the meeting was unfortunately canceled, but the community met anyway, and it turned out to be a really amazing event, in the sense that the community really came together and talked about many of the issues that you are talking about. The severe degree of fragmentation, the lack of specialty expertise, availability, challenges with insurance reimbursement, challenges with coding systems not existing. Challenges with, you know, not having the ability to actually recognize rare disease patients in the EHR and lack of access to trials. So many of those same issues really kind of came to the forefront in that, in that context, in the community, wrote a really wonderful kind of community position paper that came out of that. So it will be worth sharing with your listeners. I do think that you know, in general, there are two sets of problems. You know, one is a lack of knowledge about rare diseases, because these patients are rare. We know very little about rare diseases. They are often not studied very much on the basic research side, either. And then the second aspect is that there are patients with rare diseases, information is lacking in their health record, and information about each patient is fragmented across the landscape. And the rarity of patients also means it's hard, as you mentioned, to find information about those patients, to have enough information to sort of even see a clinical signature that's good enough to then find the next patient. And so those are two separate sets of problems that intersect, but

that, you know, we've been trying to, you know, chip away at, and I think, you know, with some of the momentum we're seeing now across the community, both with the ARPA proposal, but also more generally, in the rare disease community, I think progress is finally starting to be made. The Monarch Initiative has focused on integrating just knowledge and data about rare diseases that we don't collect in the context of the electronic health record. So a lot of rare disease patient characteristics are not things that are billable, you know, attributes. So we don't bill for having characteristics like, you know, widely spaced eyes, but that might be a very defining characteristic biologically, for being able to diagnose a patient with a specific rare disease. And so we, you know, we're really trying to sort of, and sometimes that information is in the notes, but a lot of times it isn't. And so it takes getting to a specialty clinic for someone to actually document those kinds of characteristics. So then use those characteristics to compare against the basic research data and other knowledge sources. And so there's this mismatch between really needing to think about diagnostics and characterization of patients with rare diseases from a biological perspective, which is not really what our electronic health record systems and our standard clinical encounters are really designed to do. And so bringing those two worlds together is one of the main areas of focus of our community. The second thing has to do with this, this fragmentation of electronic health record information as it stands. And I, when I teach, I often ask my students a trick question, which is, how many undiagnosed patients do we have in our electronic health record data warehouse? And of course, the answer is zero, because everyone must get a diagnosis. But it's because for billing purposes. They, you know, more generalized codes are used in order to provide care, and oftentimes, not as only are the codes more generalized, but they might be exactly wrong, because there's the clinician might know that if they choose a specific code that might get the patient the care and the insurance coverage that they need, and so we're capturing data wrongly about rare disease patients. We have estimated that approximately 80% of rare diseases don't have an ICD code, and have worked really hard with many different communities developing different rare disease terminologies and knowledge bases to build a you know, reconciled terminology of rare disease codes that can be integrated with ICD, with SNOMED, with EHR systems, and have been working with the Intelligent Medical Objects company in order to deliver that mondo terminology of rare disease codes that links back to organizations like OMIM and Orphanet and other rare disease experts so that we can actually start getting the right rare disease codes into the electronic health record. Now that does not solve the billing problem, but it would, would potentially, you know, aid the longer term strategy that you all need to find patients that have a specific rare disease and who would not be so miscoded, and they could be doubly coded. They could be coded for their billing code, that's, you know, you know, cardiac condition not otherwise specified, alongside with their very specific rare disease code. So that there's new ways of querying EHR data and coding patients that have rare diseases, and so really, kind of working at that, and to just improve the overall data that's actually in the electronic health record, in the way in which we code rare disease patients.

Dan Housman:

And back to the network problem, like, sort of the scale up of even if we could solve this, imagine at just UNC how would we solve this in the United States? I know you've had this

incredible work that you did during COVID to bring together lots of institutions, as well as the work on All of Us, which is another national initiative. How, how do you see that playing out? And, you know, maybe in the work, because I was a bit involved, but you were much more involved in the response to this ARPA-H grant. How do you think a network can play out? Is there a rapid path to a large number of health systems adopting something that will make something work for identifying patients or even just contributing data to do useful diagnostic work on these populations.

Melissa Haendel:

Well, there's definitely, you know, the philosophy of, it takes a village. And I think, you know, rare disease patients are some of the most motivated, you know, communities in the world. And so they, and they, really, you know, spend a lot of time driving around with their, you know, suitcases full of electronic health records, PDFs printed out, and have a very hard time getting access to, even to their own data. So I think they're really an important force to help make change for how this is done. You know, I think, you know, fundamentally we have very just recently, within the Center for Linkage and Acquisition of data, the CLAD for the All of Us program worked with eHealth Exchange, which is a national health information network that has allowed us to, for the first time, tap into those EHR data exchange systems that are present for care purposes, For the first time for research at a national level. And this is very exciting, because with an individual patient's consent, we can go out and get all their EHR data through those health information networks, and so that really sort of creates a much better longitudinal record. And then also with the patient's permission, we can also link it to other important data, like claims data that once we have all of that, that picture, we can start to really characterize, you know, let's say there's only 10 patients with a given rare disease in our country. But if we can get all the data about each of those 10 patients, we can really build, you know, computable phenotypes that would help identify those patients within the electronic health record and provide clinical decision support to even general practitioners to say, Hey, your patient has characteristics that look like it might be one of these ultra rare diseases, you should either refer them to this type of specialist, or here's the next lab that you should order and really just provide that kind of clinical decision support at the point of Care, no matter where that patient may be being seen for care. And, you know, often the problem with patients like that is, if they don't get their care at a, you know, more sophisticated Diagnostic Center, like at an Academic Medical Center, it often takes many years before general practitioners out in the sort of what I call the real world, you know, recognize them as potentially needing that type of care, if at all. And so we really need to be able to put the patient back together again and go out and get all the data about each patient, especially the rare disease patients, but also, you know, control patients and patients with common diseases, so that we can, you know, find the needle in the haystack and be able to find those patients that you know, folks like yourself, might need for a given clinical trial. They're out there. We're just not finding them because we don't have, you know, enough of a complete record about each one, and they're not coded in any way that's, you know, so obvious to find them.

Dan Housman:

So we at Graticule, we look like a CRO or look like a technology company with the pharma groups, we have a network of health systems already set up to do different kinds of things, directly in Epic or directly in other EHRs for either patient ID or EHR-EDC. And the kind of question I think sponsors will ask us, and so I'll ask on their behalf. Back to you and the academic world. What does a sponsor need to do to make a big network of sites? Get excited to be able to collaborate. You know? What do you think is the missing ingredient to make something happen here in rare disease? And I'm putting you on the spot, I know, but I'm thinking, What do you want to see from the life sciences, med, device, world to come towards you?

Melissa Haendel:

I mean, I think you know, for these very small populations, obviously, just finding them and having, you know, really good, you know, ascertainment and retention and really thinking about it, you know, more as a partnership with those communities, as opposed to, you know, larger trials where, you know they're just, you know, patients obviously are important, but there, there is enough of them that there's not so much special, you know, communication or special education that's needed, and so I think that's where you know, so that's part of it. And I know that's kind of a fuzzy answer, but recognizing that each patient communicate and care, odysseys that these patients go through are really hard, and recognizing that they actually have to be partners in designing the studies and partners in making sure that those patients are really sort of deeply engaged in the studies, I think they you know, those those patient groups, are incredibly powerful, as I mentioned the beginning. From the healthcare perspective, these are often patients that are hard to care for, have poor outcomes and are expensive, and so the healthcare systems are also motivated to really help these patients as well, and want to do something about them. And they need it to be easier. They need it to be easier to find the patients. They need it to be easier to interact with the patients. And they want to be able to track outcomes for any of these interventions that are going on for a trial or any other research study. And so I think, you know, also just sort of thinking about the technologies that can be deployed at the point of care that can help with that also sort of reinforces this idea of the patients as collaborators and healthcare systems as collaborators in the process of doing a rare disease trial. And I think we just have to really think about the way in which we do studies in rare disease patients a little bit differently than the way that we do with much more common diseases. There just has to be a lot greater, you know, investment in those relationships, but with great dividends. You know, this is a little bit of an analogy, but I happen to be on the advisory board for the ECHO program (Environmental influences on Child Health Outcomes), which is a national pediatric observational study. And they also do trials in those kids across the country, and it's mostly rural kids. They have a 100% retention rate in those kids, and it's because the families that have those kids are so invested in being part of research. And I think that's the kind of sentiment that we need to have with these rare disease groups, you know, once they're identified and participating in a study, that they're really deeply engaged in that way, and feel like they're, you know, kind of co-designing and co-driving the research. And that's where I think

there's not been as much experience for many pharma companies in doing that kind of partnership model.

Dan Housman:

And do you think that you, or you, plus the network that built out things like N3C and All of Us would be able to respond to we were trying to run a study? You know, we want to have 500 patients in an RCT. Can those groups help find them and use some of the existing EHR collaborations? Governance? I mean, I know how hard it is to get 100 institutions. Everyone will look at it together to solve a problem.

Melissa Haendel:

I mean, I think we, you know, there are policy changes needed nationally in order to make this easier, you know, organizations like N3C are not allowed to re-identify or re-contact patients. So that's a dead end for that kind of study. However, the All of Us Research Program is a consented study, and you know, they have been talking quite a lot about, you know, partnering with pharma for recruitment purposes, and would have the ability to do that. And I think that's a wonderful idea and that's also a really great cohort, because that cohort is, you know, not so much pediatric. They're just starting a pediatric programs, but, but otherwise it's very socio economically varied across all states. And so it's, it's a really good representative cohort. However, it's also not that big. So it's, it's, you know, their goal is to get up to a million participants. So we've got to find ways to scale beyond that for rare disease. That's not enough for rare diseases, and that's where you know these kinds of policy changes, and also use of the health information networks for research purposes could potentially come into play. There are other organizations that are also harmonizing data, like Truveta, that could also be partnered with, but ideally, you know, this type of, you know, data harmonization for the purposes of doing computable phenotyping, finding patients and recruiting them to trials, would be a public service utility, not unlike electricity. It should be, you know, government regulated, but a private industry, where we actually have the connectivity that we need. Some may say that, you know, HL7 FHIR is going to get us there, because we're going to have all this great interoperability. But what I have seen is that people can, you know, push data and various ways into FHIR, and it is improving things, but it doesn't create the policies that allow the actual querying and sharing and harmonization across different organizations that are actually needed in order to find the patients and recruit them for a trial. So there's a big leap that needs to be taken, but I think we're actually there in terms of the technologies that need to be in place. What we're not there with are the policies and the governance.

Dan Housman:

I know you have to go soon. I mean, I was gonna ask one other question about, you know, in that ARPA-H Grant, there was definitely a thread of AI-ML, and what are your thoughts, because, we at least at Graticule here quite a bit, are putting our own investments in around using the latest and greatest ways to process free text with AI, and you can see many places to

apply it. What are your thoughts about how that can help accelerate some of what we're trying to do in rare diseases?

Melissa Haendel:

Oh, I think it's an incredible opportunity. And we ourselves are doing a lot of, you know, experiments and, you know, trying to figure out new ways of getting information extracted, getting information aligned, you know it's, you know, historically, a lot of the rare disease information that you might need is locked away in the notes. And so that's, you know, that's really helpful to extract information from the notes. But it's also even more than that, it's, it's really helpful to sort of structure content together from all the different you know, places that key information is littered across the EHR in such a way that, you know, provides that clinical signature that we can use to compare and contrast patients across different contexts. And so those you know, those signatures are, you know, had been lacking, largely because of a lack of access to, you know, readily re-identifiable notes. And now we can extract that information and structure it together with the more structured information, and share that information across sites. So I think that's that's really exciting, also just the clinical decision support opportunities. So, you know, we, we've historically used what we refer to fondly as good old fashioned AI, which is our semantic inference using description logics to say, you know, if you, you know, have co-occurrence of specific phenotypes, you might want to do with this lab or this next evaluation. This is really helpful for multi system disorders, for example, you're being seen by an ophthalmologist. They don't really know to look for liver signatures, but that multi system disease might have a liver signature, right? So the the LLM approaches can really dive in and help us, you know, make those reveal those potential observations and next steps for clinical decision support much more readily apparent, not only from within one's own health system, where these events are rare and there's it's historically not been very good, but now we can start to share them better across individual sites, and then bring that clinical decision support back to the site. And so I think we're going to just see, I mean, we're in the middle of a healthcare revolution, really, with LLMs, and especially in the rare disease space.

Dan Housman:

Yeah. I mean, I do think the upshot of what you're saying is the time is now to act from the sponsor side. And you know, there's lots of opportunities to collaborate to build out these new capabilities even at a small scale. And then I think the big piece that I'd love to see, I know we've talked about it before, is just get it to scale, whether it's the government doing it or, you know, if life sciences companies have the focus, which is often the challenge, or the risk tolerance to go through some learning they can get the benefits of working with the UNC, working with, you know, networks of groups connected to All of Us to build out the things that are necessary to improve their their ability to serve patients.

Melissa Haendel:

So I wish I agree more. And I think, you know, pre competitive space, where, where we share the infrastructure to build computable phenotypes and share, you know, and share the cohorts

we share, you know, we you know, again, that public service model, that public utility model, you know, everyone should be able to tap into the electricity, even if you're using it to run your freezer in one place, and your, you know, and you have, you know, your air compressor, for your, you know, for your power washer and somewhere else, right? And so that's that kind of analogy, I think, is really where pharma companies need to go. We need to, we need to collectively build that public utility, and we need to push on our government to change the policies to let us do it.

Dan Housman:

Well, great. Well, I love talking with you all the time. Miss Melissa. I wish we had more time so this has been fun, and I hope we, for the folks who are listening, feel free to reach out with the emails, comments, thoughts, for ways to apply some of the things that are going on to your own challenges. And we were, I think both of us are just excited to see what comes next and help

Melissa Haendel:

Absolutely. It's been great being here today with you

Dan Housman:

Thanks, Melissa, catch you soon.