

Adrian Hernandez ([00:04](#)):

Hey, this is Adrian Hernandez, and welcome to the NIH Collaboratory Grand Rounds podcast. We're here to give you some extra time with our speaker and ask them the tough and interesting questions you want to hear most. If you haven't already, we hope you'll watch the full Grand Rounds webinar recording to learn more. All of our Grand Rounds content can be found at rethinkingclinicaltrials.org. Thanks for joining.

Lesley Curtis ([00:29](#)):

Today we're here with Josh Denny, who will be reflecting on advances at the intersection of digital health, electronic health records, and pragmatic clinical trials. Josh, we're so happy to have you on this podcast recording. Welcome.

Josh Denny ([00:48](#)):

Thank you very much, Lesley. It's great to be here.

Lesley Curtis ([00:52](#)):

Let me start by just asking you to reflect a little bit on your really deep experience using electronic health records for discovery. Tell us a little bit about kind of how you approach that and maybe what you've learned.

Josh Denny ([01:12](#)):

You know, it's been a long journey for me and it began really as a medical student, thinking about in the early days and electronic health record system at Vanderbilt, which was homegrown, of how you could start to take some of those observations we were making and try to systematize it. I remember just some early patients with phenomena that you realized just didn't have much literature around them, and I remember seeing an early case of a patient with triple drug therapy for HIV, and he had come in with a myocardial infarction and there was really no literature at the time of whether or not that could have any relationship to his protease inhibitors or the drugs that he was taking, but otherwise was an unusual candidate to have a myocardial infarction, and thinking it would be so powerful if we could use the electronic health records to mine for this kind of information.

Josh Denny ([02:17](#)):

And so really since late in medical school and then through my residency and faculty career, I tried to use electronic health records and find ways to tarnish that information that we collect really as a byproduct of clinical care to be a powerful tool for research, and it really has proven to be a powerful tool in a lot of settings. We've learned that everything that we thought was maybe just going to be at our fingertips and things that we thought would be easy are maybe not so easy, but there also is an incredible wealth of things that you could use electronic health records for and questions that you never thought you would be able to ask and answer that you can also dig at with electronic health records and maybe answer questions you couldn't have answered in the other way.

Lesley Curtis ([03:17](#)):

Yeah. Your point about it not being quite as easy as maybe you thought it would be. I think many of us who have worked with electronic health data in a variety of settings have that really unfortunate realization early on in the process, and it's an important one. So as you know, the collaboratory really is

focused on pragmatic clinical trials and figuring out how best to do pragmatic trials that are embedded in healthcare delivery. Really electronic health records are central to that. Can you tell us a little bit about the work that you've done with natural language processing and how that has potential really to benefit clinical trials, and certainly any experience you've had doing that kind of work?

Josh Denny ([04:18](#)):

I actually got started with electronic health records through natural language processing or NLP, which is perhaps surprising. I think most people in most environments started with more structured information, especially billing codes. One of my early projects was looking at a rare physical exam finding and looking at its impact on mortality. What we started with as a premise that you could take electronic health records and you could take their natural narrative content and really try to understand the course of the diseases and treatments that an individual has. When we were starting this, there wasn't a good record of medications an individual received in a structured format. So the only way to get that was through NLP, and we found that we could provide a good record of medication exposure and dosages using NLP and use that for, for instance, pharmacogenetic studies and find things that we're supposed to find.

Josh Denny ([05:24](#)):

A lot of that still required manual curation, but it would reduce the hypothesis space from everyone exposed or everyone that might have a certain outcome to a much smaller data set. In some cases we could do it entirely through automated means, such as with warfarin pharmacogenetics. We were able to build our warfarin dose response really entirely using NLP because we got it accurate enough. It wasn't perfect, but it was accurate enough that over lots of people, you can define really good understandings of dosages and INRs that individuals had on warfarin and then associated that with genetic changes that would predict that relationship.

Lesley Curtis ([06:12](#)):

That's a great example, Josh, and talk a little bit, if you would, about how those approaches translate across institutions. What do we know about that?

Josh Denny ([06:30](#)):

We've been building, through part of the electronic medical records and genomics network, phenotype algorithms across multiple sites for I guess a little over 10 years now, and when we started that, we believed that you could probably build the ... we thought building such an algorithm and transporting it would be easier often than it was. You find, even with billing codes, some of those billing codes would be different in how they're used at different sites, and they're used differently at different times. Some of the elements, like looking for exposure to contrast media in radiology reports, that kind of stuff actually transports really pretty well, and so what we found is we've been able to take a lot of algorithms, and if you think about how they work in a couple of different sites, a lot of times they seem to transport pretty well.

Josh Denny ([07:36](#)):

If you take our type two diabetes algorithm that we developed in Emerge. When we looked at this I think in 2016 or so, we found that about 40 other sites had adopted and used it within their healthcare systems, and for the most part, the system just relied on structured elements, medications, lab results, billing codes, and transported pretty well. One of the things we've found that has been really challenging for people to implement across a number of sites has been natural language processing. If

you think about algorithms coming into either being a combination of Boolean logics, like and, or, not, versus machine learning approaches, which are becoming increasingly interesting and powerful, we find that the and, or, not kind of logic, the Boolean logic based algorithm and deterministic logics, maybe transport better than those built on machine learning.

Josh Denny ([08:36](#)):

It's not universal, but it has to do a lot with kind of data once it came in. One example I could talk about machine learning is one of the algorithms we built early on in this process was finding patients that had colorectal cancer, and when we really investigated the model, it was getting just amazing performance in our data set, really honestly probably too good to be true, is it was learning some of the features based on the formatting of the vital signs, and obviously that's something that doesn't transport to necessarily another site, nor is it stable within our own EHR, and so you have to make sure that those kinds of algorithms, before you think about transporting them, are of course stable within your environment and that you try to minimize the ways in which they learn features that are not actually medically or biologically relevant to predicting the disease or a phenotype we're looking for.

Lesley Curtis ([09:37](#)):

Yeah, that's really important learnings, and again, as you said, probably early on, many of us who've done work in this area think it should be easier than it actually is, and then we're reminded of the complexity in the system that that makes using any kind of data more challenging than we might expect. Josh, I was certainly very excited to hear of your move to the All Of Us research program, and we'd love to shift the conversation to hear more about All Of Us and the work that you're doing there.

Josh Denny ([10:20](#)):

Certainly. Our family and I moved up here in January of 2020. It's been an eventful last five months, for our nation, for the world. I've been involved with All Of Us, though, since really the end of 2014 in some fashion, and so joining it from the NIH perspective was ... there's no such thing as maybe an easy move, but it was a very natural move and one I've been very excited about. I think a lot of the ... All Of Us is built on some of the kinds of works we've done in other networks and just trying to scale them, and really getting at some of the ideas that you have behind the NIH collaboratory in building a population that is useful for many different kinds of research by coming in, by being re contactable, by sharing electronic health record data, as well as survey data, bio specimens, and looking towards the future of linking in other types of information, as well as we have some digital health technology information coming in now.

Josh Denny ([11:31](#)):

You build a cohort that could be useful for discovery research, for clinical trials, for types of information that you get from the health record, as well as types of information you get that are patient reported outcomes, and we can look at those exposures, those outcomes all over time. So we've been very excited to become part of this from the NIH side and think of how we can help propel research going forward.

Lesley Curtis ([12:00](#)):

And you mentioned it has been quite a few several months since you arrived with the NIH. Can you at least touch a little bit on how you see all of us helping to address the pandemic, the COVID-19 pandemic?

Josh Denny ([12:20](#)):

Certainly. So All Of Us is first and foremost a platform. We partner with our participants to produce a research platform that many others will use, and so I'll just say briefly that what we call the research or work bench by which people can come in and do research on the platform is now available in its beta launch, and so academic researchers at US institutions right now in this beta forum can come in and start using the resource. That being said, most of the data, or actually all the data that are out there right now are pre COVID-19, and so since COVID-19, we've, like many other research project, had to adjust, and part of that, the first element of that is safety and what we need to do to respond, and so we paused all of our in person activities in March.

Josh Denny ([13:20](#)):

And then the next element of that is how can we help? And I decided to pursue a couple elements. One is looking at antibody prevalence in individuals who enrolled March 2020 and before. So we've had about 3000 people contributing biospecimens each week, and as we look back in time, we should be able to get some evidence of where and when the virus came into the United States and which populations it seemed to infect with our ... we have a very diverse population with over 50% racial and ethnic minorities, and over 80% being underrepresented populations by some metric, whether it be socioeconomic or education, rural location versus more common, or sexual gender minorities, and so the antibody tests are one of those. The other two are we decided to launch a survey that investigates the influences of COVID-19 on health more broadly. So mental health, social distancing, of course COVID-19 symptoms and diagnoses, and then also looking at the effects of things like effects on optimism, resilience, depression, anxiety, stress, economic impacts, and discrimination, which of course is front and center with this country right now, as we're in June and the recent events with George Floyd amongst such a long history, but really brought back to the front of our minds with what's happening now.

Josh Denny ([14:58](#)):

So that's a repeated survey that people are answering starting in May and we'll run through the summer, and then the final thing is we're trying to promote electronic health records standards for reporting of COVID-19 diagnoses and testing to make that resource for researchers robust to investigations around COVID-19 and lasting impacts. We won't be the tool that allows people to really track diagnoses on a kind of day to day basis. We're building a resource that is probably more relevant for longterm outcomes and effects and potentially things like being re contactable for future clinical trials or things like that could also be part of that. So we're trying to build up that resource that other researchers can come in and do those sorts of investigations, and of course we're working with CDC and many others on these projects.

Lesley Curtis ([15:53](#)):

Josh it's really exciting and really gratifying to hear how this platform is already being put to what should be really impactful use. So I'm really glad to hear more about that. I do want to, again, thank you for joining me today. This has been really a great several minutes with you, and I encourage our listeners to join us for our next podcast, as we continue to highlight fascinating and informative changes in the research world. Again, thanks very much for joining us today, Josh.

Josh Denny ([16:37](#)):

And thank you so much, Lesley. You take care.

Adrian Hernandez ([16:44](#)):

Thanks for joining today's NIH Collaboratory Grand Rounds podcast. Let us know what you think by rating this interview on our website, and we hope to see you again on our next Grand Rounds, Fridays at 1:00 PM Eastern time.