

Adrian H.: [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](https://rethinkingclinicaltrials.org). Thanks for joining.

Adrian H.: [00:28] So, I wanna welcome everyone to today's podcast. Today we're here with George Hripcsak and he just finished a great presentation on OHDSI: Drawing Reproducible Conclusions from Observational Clinical Data. And so, from that, we'd like to extend the conversation in terms of learning more about OHDSI and it's future. So, George, thanks for joining us.

George H.: [00:49] Thank you for having me.

Adrian H.: [00:51] You know, it's pretty fascinating about how this developed. Can you give us a little bit of more history of how OHDSI started and why?

George H.: [00:58] Okay, very good. Well, first of all, many of us in OHDSI started in the field 30 years ago, at Columbia in 1988. And we've been working on observational health data, all though we weren't calling it quite the same thing back then. OHDSI started as OMOP, Observational Medical Outcomes Partnership, about, let's see, eight or nine years ago, now. And it was funded by pharmaceutical industry to develop methods to improve observational research. And it was funded from pharmaceutical industry to the NIH Foundation to several academic researchers to build methodologies. So, for five years they worked hard to come up with methods to carry out observational research better. So, they developed a data model, terminology, vocabularies, that is, and statistical methods that would produce more reliable evidence. And frankly, during that time they were building a team.

George H.: [01:56] That project ended naturally at the end of its five year term, and that's when it took a turn where, under the original agreement they were not allowed to do clinical research, that is, they could only develop methods, not carry them out to answer clinical questions. So at that point, the researchers from OMOP separated off, outside of the government, and formed OHDSI with Colombia being the coordinating center, but with participants now from around the world, 25 countries, who can continue to develop methods and data models, but also can actually carry out clinical research. And that's been going on since the end of 2014.

Adrian H.: [02:33] And that development, what were some of the challenges? Was it coming to standards around common data model, governance, or the methodology around observational research here?

George H.: [02:47] The data model ended up being a huge challenge, because each database is in a different format, and as you look around the world certainly, each one is in a different ... uses a different vocabulary, but even within the US, many different vocabularies are used. So, it was a large effort to create the mappings from all of these vocabularies to a single common vocabulary or set of vocabularies, I should say. And the common data

model evolved forward, so that was one part. The other big thing, as you alluded to, was the observational research methods. And so, not the big surprise I'll say, but the big recognition is just how much we underestimate how uncertain our research results from observational data are. In other words, we'll write a paper that says that this is the answer I got, and here's the confidence interval around that result, and as the databases get bigger, our confidence intervals get smaller. But then if you do and you try to verify that using some of the methods that I talked about earlier, from OHDSI and from outside of OHDSI frankly, you see that these confidence intervals need to be...

George H.: [03:58] There's much more uncertainty than is relayed in our confidence intervals and that it's causing a problem and so then we come up with new methods. There are self-controlled case series and self-controlled cohort studies that try to have a patient serve as his own control. There are these calibration methods that look at negative controls to try to figure out what's going on, propensity score matching and so forth. So, a lot of methods are needed and we're not done coming up with them yet, obviously.

Adrian H.: [04:27] And now kinda turning to a bit, what are the current activities? What's been the focus and or examples of projects or programs with OHDSI?

George H.: [04:39] So, what we've been doing, is each year we come up with one major challenge to address at our annual symposium, which is held in October or September. And in the previous years ... The first year it was tools and getting the data model shared more broadly. The second year was much of what I showed today on doing the large-scale evidence. And the third year was what I showed last, which was Howoften.org, that is, for every drug available around the world and every side effect coded in SNOMED, which is 300 thousand terms, what is the rate at which these things occur? For this next year, what we're trying to do mainly and it's not decided exactly what will be the big fanfare in October yet, but it's extending beyond the depression study that we did as a proof of concept to actually generating the evidence across many clinical areas. So, basically scaling it up and actually doing the job that we, so far just did as a proof of concept.

Adrian H.: [05:40] Now, you mentioned kinda the openness of this. How do people get involved? The symposium in the fall, sounds like things have been really spreading out.

George H.: [05:51] Well, our fall symposium is held always in Washington D.C. We do that so that our government collaborators can join the conference more easily and we can get more government employees from NIH, from the FDA, and other agencies can easily join us. Our conference so far, and we have ... Last time we had about 450 people is always free to everybody. So, it's a completely free conference with almost 500 people and free lunch for everybody. So, we need to raise money to do that, so we increase our availability through not charging fees and not making money on the deal. Then of course, as I explained, all our software goes on the internet. As we do studies, we try to release them on the internet as quickly as possible. When I go and do talks, I'm used to people asking me, let's say a panel, each member of a panel does a talk, and then they get asked "Well, can we share your slides? Are there any private things?" And so nowadays, I can always say "Do whatever you want, it's probably already on the internet."

George H.: [06:58] So, I no longer have to worry about hiding my stuff because I've already shared it. And sometimes it's embarrassing 'cause we don't always have good results or the results we expected, yet we share it all the time.

Adrian H.: [07:08] Well, that's really impressive, the openness and sharing is really a great set of values here for, not only OHDSI, but also for everyone in research. What do you see in say, three to five years? How do you see OHDSI positioned and what will it be doing?

George H.: [07:27] Well, number one, we wanna expand the network to even more countries. We're doing pretty well, but we wanna continue. South Korea will be moving, I believe to the entire country. Germany is going forward with a large nationwide data network, also. So, we hope to see more of that. We wanna extend better into South America and Africa, and so we cover, really, the whole world population. We did temporarily, almost have Antarctica covered, but we hope to someday have seven continents covered by OHDSI. On methodology, we, of course, expect to do further research in what are the best ways to carry this out and to test our assumptions that we're making.

George H.: [08:15] And frankly, we wanna be generating large-scale evidence and having people use it day in and day out. An area I didn't talk about earlier, was patient level prediction, that is, given a specific patient, with specific properties, what are that person's risk as opposed to the overall risk? And we have a work group that's been doing a good job on that, but that's something I didn't talk about and I think will be, basically the basis of precision medicine and we hope to in the future be working further on that.

Adrian H.: [08:46] And one of the other kind of areas that people are talking about often is linkage. Do you see that as a key component for OHDSI or how do you see that happening in the future?

George H.: [09:00] So, first of all, OHDSI never links any databases 'cause they don't actually hold the database itself. The OHDSI members hold the database behind their firewall and insofar as they can link their database to say, say they have an electronic health record database, if they can link it to a paired database, that would make that database stronger and more valuable in studies. But we can't do the linking 'cause we don't have access to the patient identities. I think that certain projects will go forward and will do some of that linking. The kind of linking we do now, is what I spoke, perhaps about earlier, which was, say linking the tumor registry to our electronic health record database.

George H.: [09:43] So, we're doing that as we go forward. The All of Us Research Program is in fact, using the OHDSI data model as the basis of the clinical data and we're aligning that with the genomic results that we will be getting in the future. And that is probably gonna be the first test bit of doing that in OHDSI. Furthermore, the eMERGE network, which is NIH led consortium, which compares electronic health record data to genomic data, has adopted the same OHDSI data model. And that will be another pilot where we start figuring out, well how do we put these two things together? So, that would also be in the five year horizon.

Adrian H.: [10:18] Well, that's terrific. So, it sounds like a ton going on now and a ton going on in the future. So, really a wealth of data that's now, it sounds like really being used for generating evidence and it sounds like soon you'll be able to expand and cover all seven continents. That's great. So, I wanna thank you for joining us on this podcast and also thank everyone else who are listening and then as people consider joining us again, our next podcast will be with Dr. Rachel Florence and launching the NEST Coordinating Center Data Network to improve the use of real-world evidence in the medical device ecosystem. Hope everyone can join us then.

Adrian H.: [11:09] 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 p.m. Eastern time.