NIH Collaboratory Podcast 22: Development of Harmonized Outcome Measures for Use in Research and Clinical Practice

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 some of 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.
Adrian H.:	00:27	Hi, this is Adrian Hernandez and I'm a moderator for the NIH Collaboratory Grand Rounds, and today we have on our podcast a discussion around a recent Grand Rounds on the development of harmonized outcome measures for use in research and clinical practice. We have Elise Berliner and Michelle Leavy, who have been leading those efforts along with others, and we want to learn a little more about what they did and what problems they've been solving.
Adrian H.:	00:54	So, Elise, let me start with you. This has certainly been a major interest in HRQ for many years. What's the problem that you all have been trying to solve here?
Elise Berliner:	01:05	So I've been working in systematic review for more than 19 years, and we always have a problem that different studies use different outcome measures. It's impossible to synthesize across studies. One topic in particular that we've been working on with our colleagues at Medicare is on treatment-resistant depression and new devices for that. We actually did a review trying to figure out what the definition of treatment-resistant depression is and how it's defined in different studies, and we found that there's huge variation. So, this creates all sorts of downstream issues for CMS. If they want to have a coverage policy, then what do they use as the definition? So those are the kind of problems we've been working on for a really long time, and so this project came out of that.
Adrian H.:	02:02	That's certainly interesting because, certainly, it's an important area and a common example where you need to understand the outcome that you're trying to change over time. And if you're not going to have any agreement on the outcome, it can be quite difficult to know whether things are improving or not.
Adrian H.:	02:19	Well, Michelle, you're a veteran in the area. You've been working in this area for a long time. How did you all approach the problem put forth? What was the strategy here?
Michelle Leavy:	02:31	Well, we started with, several years ago, the conceptual work that laid the foundation for this project. We worked with quite a

		number of stakeholders to build the outcome measures framework. We went to registries in different condition areas to understand what outcomes they captured and then to work through an idea of how you could classify those consistently in a way that made sense across the different condition areas. That work led us to the outcome measures framework, and then using the framework, we were able to really tackle the problem of harmonizing outcomes in these specific clinical areas. It gave us a useful tool to take all of the outcomes that were found... In AFib, for example, we identified I think 113 different outcomes, and it gave us a way of organizing them and talking about them with the registries so that we could start to think about how we might come to a minimum set and how we would want to then agree on harmonized definitions for each of those.
Adrian H.:	03:33	Michelle, can you talk a little more about the process? Because when you describe stakeholders, sometimes that can mean one or two people or a lot more. What was the approach there in terms of their engagement?
Michelle Leavy:	03:46	We worked with work groups in each of the clinical areas. Our core group was made up of registries. We had on average about 10 registries in each group. We tried to identify as many registries as we could that were currently collecting data in that area and that were focused on patient outcomes. We invited them to participate and then we rounded out that group with a broader set of stakeholders who brought in perspectives from health systems, from other research organizations. We tried to incorporate payer perspectives in that group, patient perspectives. And in cases where they weren't represented by the registries, we also tried to bring in representatives from pharmaceutical and manufacturing as well. Within each condition area, we averaged about 20 people in the work group altogether.
Adrian H.:	04:47	Wow, that's impressive. Tell me a little bit more about what's the end product look like? What's the ultimate outcome here that you all are generating?
Michelle Leavy:	04:58	So for each clinical area, we developed a minimum set of outcome measures. The idea with the minimum set was that these would be measures that could be collected in routine clinical practice, they'd be suitable for use in registries, and wherever possible, they would connect into existing widely-used measures, like quality measures from CMS, for example. For each measure within the minimum set, we agreed on a harmonized narrative definition of the measure, and then we took that narrative definition and translated it into standardized

terminologies wherever we could. We tried to get into as much detail as we could so that we could capture these in existing data sets. So, we defined clear timeframes of interest and wherever possible tried to use, for example, in patient-reported outcomes, we tried to get to the level of recommending a specific validated instrument when we could.

Adrian H.: 06:01 Okay, great. And, as you know, based on this project and other work, one of the worries is adding more to the clinician's plate. There's a lot of data out there that their days are being extended and people are worried about them becoming more and more data entry folks and adding more may be difficult, even though the goals are noble. How was that considered in terms of the harmonization and outcome measure development in terms of what you described, trying to get into routine care and thinking about that?

Michelle Leavy: 06:43 Yeah, I would say burden was one of the major considerations throughout the entire process. We were trying not to add anything that is not already routinely captured, but rather to take what would be routinely captured and agree on a standardized way to capture it with the idea that from a big picture standpoint, if more of these data collection efforts were aligned, there would be less need to have duplicate data capturing, and hopefully, the overall burden of data collection would come down. But it was a sticking point across the work groups, particularly when we were thinking about things like long-term follow-up of patients and how often we might want to capture some of this information.

Elise Berliner: 07:29 I think one of the really interesting things is that we did have a broad stakeholder group. So, there were some people there who were really focused on the clinical feasibility, and then there were researchers who really were focused on getting complicated validated outcome measures. And we were trying to strike a balance between that. One particular example of that I think was - again for the depression example - how often should you collect the information and how tight does the window have to be?

Elise Berliner: 08:12 So we had someone from Minnesota Community Measurement, which is a project to collect data across the whole state of Minnesota, and the person who was representing that project had a lot of experience trying to get the data and was saying that it was totally not feasible to get everybody at a six-month time frame and maybe at six months plus or minus two months. And the researchers were saying, "Well, that's not valid for my research project." So, we were really trying to balance that, and

		I think that that is something that, going forward, it's a question. We want to use real world data, but there are practicalities with collecting it.
Adrian H.:	08:57	Very true. The other component is incorporating patient perspectives. Can you talk a little more about how that was done and were you able to get to consensus?
Michelle Leavy:	09:10	In going into the project, we wanted to incorporate patient perspectives in each of the groups. We brought in patient representatives as part of our stakeholder groups in each of the different clinical areas. So, they did give feedback on the measures that were included in the minimum set and on the definition. But where we found it extremely difficult was our work did not extend to developing any new instruments. It was really reflecting what already existed, and in some cases, we couldn't reach consensus on how to capture some of the things that were of interest to patients.
Michelle Leavy:	09:54	In some cases, there just weren't really instruments available to capture things. Like in asthma, we talked a lot about missed days for school and missed days for work both for caregivers and for patients, and we didn't have any good ways of capturing that. Then in other cases, they were just dependent on the different needs of different patients. There wasn't a single instrument that we could recommend, but rather... In depression, for example, we heard that what would be of interest to patients with more severe depression may not be the same as what would be of interest to patients with milder forms of depression. It was a really difficult area.
Elise Berliner:	10:35	I thought one of the fascinating things also was how much it varied between the different topics. For atrial fibrillation, it seemed like the clinicians hadn't really thought at all about using any kind of patient-reported outcomes. So, there were a few that existed, but people didn't really have experience with them.
Elise Berliner:	10:58	For lung cancer, we were looking at a lot of different kinds of treatments, including surgical treatment and chemotherapy, and the patient reps there really almost couldn't say which were the prioritized, most important things to patients. So that was really interesting too. I don't know. Maybe, Michelle, you could talk more about that?
Michelle Leavy:	11:29	Yeah, in lung cancer, we were looking at non-small-cell lung cancer broadly and we found that what was of interest in some ways varied depending on the type and intent of treatment. So

		again, it was an example where what you might ask someone in an earlier stage was really different from what you might ask someone with a much later stage of the disease. So, lung cancer was an area where we only came to consensus on some domains that might be relevant to measure in terms of patient-reported outcomes, but we couldn't get to the level of a particular instrument that might be useful. We also found in that area that a lot of the instruments that our registries and our researchers were familiar with were quite long and maybe not as feasible to use in routine clinical practice.
Adrian H.:	12:25	Certainly, it's been impressive to see what has come together. Elise, what's next? How do you see, now that this has been done, what's next from this project and program?
Elise Berliner:	12:37	So for now, we're trying to tell as many people as possible about what we've done and get feedback and think about implementation projects. In particular, we do have some additional funding from the Office of the Secretary of Patient-Centered Outcomes Research Trust Fund for a capstone follow-on project. That project, right now we're in the contract negotiation stage, but the statement of work is public on FedBizOpps. So I can tell you that we're going to actually go into clinical sites and work with clinical registries and really try and implement these outcome measures in the depression field, and we're going to try and set up the infrastructure for collecting patient-reported outcomes and the other outcomes in clinical sites, exchange the data with registries, and set up the infrastructure to do research projects. We want to work with both primary care and with psychiatric specialists. One particular research question of interest is coordinated care between primary care and specialist care. So, we'll both be testing out the burden and the feasibility; but hopefully, we'll be able to get over those humps and really set up an infrastructure for doing some good research projects.
Adrian H.:	14:09	Terrific. It's really good to hear that you're not stopping just with this, that it's really towards implementation, refinement and scaling. So, great.
Adrian H.:	14:21	So Elise and Michelle, thanks for spending time with us on this podcast. I hope everyone enjoyed it, and please join us for our next podcast as we continue to highlight fascinating and informative changes in the research world.
Adrian H.:	14:36	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.