Jill Harrison:

Hi, this is Jill Harrison, Executive Director of the National Institute on Aging IMPACT Collaboratory at Brown University. Welcome to the IMPACT Collaboratory Grand Rounds Podcast. We're here to give you some extra time with our speakers and ask them the interesting questions that 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 the companion Grand Rounds content can be found at impactcollaboratory.org. Thanks for joining.

Hello everyone. I'm Jill Harrison, one of the Executive Directors at the NIA IMPACT Collaboratory. I'm joined today by Dr. Julie Bynum. Dr. Bynum is the Margaret Terpenning Collegiate Professor of Internal Medicine in the Division of Geriatric and Palliative Medicine at the University of Michigan. She's also Vice Chair for Faculty Affairs and the Geriatric Center Associate Director for Health Policy and Research. Dr. Bynum is also the leader of the technical and data core at the NIA IMPACT Collaboratory. Dr. Bynum, thank you so much for joining me today.

Dr. Julie Bynum:

Thank you for inviting me.

Jill Harrison:

The IMPACT Collaboratory hosts a monthly Grand Rounds series, each with a companion podcast. This past week, you presented a well-attended Grand Rounds on the topic of healthcare-generated data to identify people living with dementia for embedded pragmatic trials. Can you please explain what healthcare-generated data is and what it has to do in terms of conducting dementia research, specifically in pragmatic clinical trials?

Dr. Julie Bynum:

Yes, I'm very happy to do that. Healthcare-generated data is a term that, basically, we in the technical data core created when we were writing our first description of our activities. What we mean by healthcare-generated data is any systematically collected information, like billing data or electronic health data, that's generated through the process of delivering care. The advantage of using that kind of data is it's already part of the workflow, so when you're doing a pragmatic trial that you want to be as close to real world and highly scalable, it's very useful to tap into data streams that are going to happen anyway, whether your research happens or not. Your second question was why that's important for pragmatic trials. Well, I just basically alluded to it that really we are looking for scalability and real world implementation, and using these healthcare-generated data make that possible, but of course, all data sources have their own strengths and weaknesses, and they're really important to consider the trade-offs of using this kind of data versus other very detailed data that's often collected for research.

Jill Harrison:

Thank you. You know, it struck me during your talk just how challenging identifying people living with dementia can be. Can you describe some of those challenges in more detail and how researchers can overcome them using types of healthcare-generated data.

Dr. Julie Bynum:

Right, yeah. Well, it all starts with how well we actually identify people living with dementia in our healthcare environments. Putting the data aside, the data use, the healthcare-generated data use aside, how well do people come forward seeking care, seeking a diagnosis, and how responsive is our

healthcare system in its ability to get an individual concerned about memory complaints to a provider who has the skills and abilities to make an accurate diagnosis, and then ultimately, what that results in, is the documentation of a diagnosis, whether in billing or in electronic health records on problem lists and diagnosis lists, and then lastly, the actual disclosure of that diagnosis to the person and the family, so that if we ask them, they can tell us whether they have the disease. So, the literature suggests that about 50% of cases, people who are actually living with dementia are not diagnosed at any given point in time.

So that alone creates all sorts of challenges for us to be able to identify these people to participate in our pragmatic trials. On top of that, some of that lack of recognition is more prevalent in groups who are underrepresented or have low literacy, for example. So we have inherent problems, challenges in just care-seeking and case identification in healthcare, and that of course gets reflected in our electronic health record data and in the administrative data that we're using.

So I think where the healthcare-generated data can help is that it actually requires people to seek care, but they don't necessarily need to seek care for their cognitive complaints, so we can pick up diagnoses in all sorts of different places. We can even pick them up, not just diagnoses, but symptoms, in the text notes. As clinicians write their notes, they allude to symptoms and complaints and relying on caregivers, for example, that can be detected with newer methods of using natural language processing and other techniques. So there are some opportunities to capture undetected cases when using electronic health record data, and the other advantage is that often with these kind of data, we have everyone, meaning everybody who's seeking care, not just people seeking care for dementia, so you can actually get a sense for who you might be missing if you're using clever techniques.

Jill Harrison:

That brings me to a question that we had from one of our listeners during your Grand Rounds. There were so many questions that we weren't able to get to all of them live, so I'd like to follow up on one here, and the question is, "How does the positive predictive value of your algorithms for dementia compare to other diseases related to aging?"

Dr. Julie Bynum:

It's actually pretty similar; there's very wide variation in the algorithms. A great example is congestive heart failure. It's one of the ones that's actually hardest to capture in billing data, which I think surprises people. It's also hard to identify in electronic health records because the ejection fraction is something that's not usually a standard structured data element, so this is a common problem across many, many diseases. Another example is it's known that for diabetes, which is clearly a disease that doesn't go away, someone might have a claim for it in one year but not the subsequent year, so this is a common problem. The trick and challenge for dementia in particular is there is no blood test or diagnostic measure that can go secondarily to back up the clinical diagnosis that we are seeking, so I think that's why it's been harder for people to lay their hands on how good is this diagnosis and feel confident, combined with the under-diagnosis in the clinical practice. But it's actually a very similar problem across many, many, many diseases.

Those of us who work with administrative data through a lot, accept that often we have to think very carefully about what we can say and what we can't say when we're using this kind of data, and that's why during my lecture, I tried to make a strong point that thinking very carefully about what your aims are, what your goal is, and how that aligns with the data that you can see is a very important step in design.

For example, if you really are seeking people with late-stage disease, you have to understand that in claims alone, we can't see whether the stage is late or early, so really that term of "fitness for use" is all about that positive predictive value. How important is it for you to get precisely a certain type of person or is level of uncertainty around that value acceptable? The other thing to remember is that predictive values are really dependent on underlying prevalence, so if you... The study that we did, both of the studies, the underlying population dementia rate was 12%, but if you are working in, say, home care or a nursing home, that underlying percent is going to be much higher, so your positive predictive value will also be much higher.

Jill Harrison:

Do you have any case studies or examples, innovations, that you've seen from healthcare systems and researchers to use this data, the billing, EHR, electronic medical records claims data, to conduct dementia research that you would like to call out for our listeners?

Dr. Julie Bynum:

Yeah. There's a lot of innovative work whereby people are trying to use other markers in the electronic health record to indicate that a person may have the dementia. I think the eRADAR algorithm is one of those that uses other factors that says, "If all these things are present, there's a higher likelihood that this person has dementia." Oftentimes, those require a secondary check or some sort of chart review or something else to make sure that you're actually capturing those diseases, because again, they're based on probability, not on the presence of indicators of a diagnosis. People are doing all sorts of creative things with trying to get at the text notes, trying to figure out severity or whether there's caregiver presence, going really deep into the text notes.

The important thing to keep in mind that we're really beginning to dig in on in the Technical Data Core is that anytime you develop something like that, that's based on the text or details of a health record, you have to keep in mind that the way people use language and use a healthcare system varies from hospital to hospital, clinic to clinic, and region to region. So when you develop those kinds of tools, they really need to be evaluated in your own setting when you set out to use them. You may not need a full deep validation, but some level of checking that you're capturing that the algorithms and those tools function similarly in your setting compared to the setting where it was developed. The field moves quickly. There's a lot out there, and what's really interesting about trying to capture the innovation is it's happening in so many fields. You can't just look at the Alzheimer's literature or just the health services research literature. You actually have to look at computer science and bioinformatics to capture all of this new literature, because it's happening all over the field, which is very exciting.

Jill Harrison:

For researchers that are looking to build their competencies in designing dementia trials using healthcare-generated data, where should they start? What types of training resources would you recommend?

Dr. Julie Bynum:

Well, the first thing I would say is always look to what already exists. Building your own fresh is always going to slow you down and really limit your ability to translate, because a lot of work has already been done, whether it's using administrative data or using EHR data, there are common data models and standards out there that already exist, so they should start there. Again, one of the things that we're... A new project that we're starting in the Technical Data Core is doing an inventory of the standard models

that are out there and how they relate to each other and what it takes to be able to implement them, thinking about what's your setting and what their data capabilities already are.

I think many of us are thinking hospitals and clinics, but today, a lot of the pragmatic trials are, we hope, going to happen in community settings, and they have other kinds of data systems. So think about what's available already in the workflow. Have an open mind to that. Those are the two places where I would start. We are developing enough materials, you can see enough examples now that we're in multiple years of the Collaboratory where you could look to the website and see what other trialists are doing to identify their populations.

Jill Harrison:

My final question for you today, you know, some of this healthcare-generated data requires a partnership in order to understand what's in the universe, what data is available, for example. Could you describe a little bit those cases when a partnership between a researcher and healthcare system is needed and what you would recommend for folks gaining entrée who are interested in using healthcare-generated data in their trial?

Dr. Julie Bynum:

Yes, that's a great question. My personal research history is that I've always used research data available through the federal government, through CMS, to get administrative data. That has a lengthy process to get approval, but actually does not require necessarily working directly with the health systems that you're working in. The moment you want to use electronic health record data or any of the internal data means developing a relationship with the settings where you plan to conduct your trial and being open to thinking very honestly with them about their readiness and capability for engaging with you. These health systems, their first priority is delivering care. They want to improve it, they want to participate in research, but they also have other priorities, so it is critically important to approach health systems and assess their readiness to engage in doing this kind of work and their capabilities to do so.

Certainly places that have a track record of having done this kind of work are good places to start, but again, remembering we're really seeking to go out broader. This is about real world and how do we get to other settings, so thinking more broadly about building your data approach in a way that is most stakeholder-friendly is an important consideration as you move forward.

Jill Harrison:

Dr. Julie Bynum, from the University of Michigan, thank you so much for sharing your work with us and all you do to improve the lives of people living with dementia. Thank you so much for your time today.

Dr. Julie Bynum:

Thank you.

Jill Harrison:

Thank you for listening to today's IMPACT Collaboratory Grand Rounds Podcast. Please be on the lookout for our next Grand Rounds and podcast next month.