

[MUSIC PLAYING]

From the campus of Harvard Medical School, this is Think Research, a podcast devoted to the stories behind clinical research. I'm Oby.

And I'm Brendan. And we are your hosts. Think Research is brought to you by Harvard Catalyst, Harvard University's clinical and translational science center.

And by NCATS, the National Center for Advancing Translational Sciences.

[MUSIC PLAYING]

Appropriate specialized care that addresses physical and mental health is a human right across all demographics. In every walk of life, sexual and gender minorities exist intersectionally with other demographic variables. Health disparities arise when language around sexual and gender minorities is not standardized, or structures do not exist to collect necessary data.

On this episode of our Community Engaged series, we're joined by researcher Harry Reyes, Dr. Li Zhou, and Rodney Vanderwarker to talk through the important research Reyes is conducting to improve the documentation of sexual orientation and gender identity in electronic health records. Harry Reyes is a PhD student in the Department of Biomedical Informatics at Columbia University. Dr. Li Zhou is an assistant professor at Harvard Medical School, and the lead investigator at the Division of General Internal Medicine and Primary Care at Brigham and Women's Hospital, and a senior medical informatician at Partners HealthCare. Rodney Vanderwarker is the executive director of adult internal medicine at Atrius Health. Thank you all for joining us. Welcome to the podcast. Harry, you recently were awarded a Pilot Grant. Could you tell us what the focus of your research is?

Sure. So I'm a predoctorate research fellow in the Medical Text Extraction, Reasoning and Mapping System Lab, or MERMS Lab, led by Dr. Zhou. And generally speaking, the goal of my research is to leverage artificial intelligence, namely machine learning and natural language processing, to promote health equity, to improve the quality and safety of patient care, and to support knowledge discovery, such as the identification of new disease subtypes. I'm also really interested in the fairness, , accountability and transparency of artificial intelligence in health care.

So tell us a little bit about how the project for which you got the Pilot Grant, how it advances those goals.

Sure so--

Maybe if you could just describe the project a little bit.

Great. So to give you some background, basically, in 2010, the National Academy of Medicine, then known as the Institute of Medicine, systematically examined the state of LGBTQ health. And what they concluded was that sexual and gender minority patients, or research, rather, to date is woefully inadequate. And they had several key recommendations, including improving data collection on sexual orientation and gender identity, or SOGI demographics, in research studies, national surveys, and electronic health records. Sexual and gender minorities are believed to exist among all ages, races, religions, and social classes, yet unlike other demographic variables, sexual orientation and gender identity, or SOGI demographics, are not routinely and uniformly collected in electronic health records, or EHRs.

This SOGI data collection gap makes it really difficult to one, estimate the number of LGBTQ or sexual and gender minorities in the population. And it also hinders the ability of the health care and research communities to assess and address the needs of LGBTQ people. And so the goal of this study was to examine documentation of sexual orientation and gender identity within Partners HealthCare, an affiliate of Harvard Medical School, to better describe that documentation and improve it.

And so we had a few aims. Specifically, we focused on primary care providers. So our lab sort of sits within the Division of General Internal Medicine and Primary Care. So a lot of work we do concerns primary care. And one of the questions we had was, are primary care providers documenting SOGI, sexual orientation and gender identity, in the electronic health record? A few years ago, just prior to the Catalyst Grant, several colleagues and I lobbied Partners HealthCare to create new fields within our electronic health record that are predefined, and allow us to collect these data in a structured format.

To their credit, they very quickly approved our request. There was a bit of a campaign, but they approved our request to incorporate these new fields for sexual orientation, gender identity, and assigned sex at birth. So gender identity needs to be compared with sex assigned at birth in order to identify patients who might be transgender in order to then facilitate research, and quality improvement, and quality assurance work within our health system.

And how long have they had this structured data in their EHR?

Only since May 2015. So Partners HealthCare had this multi-million, multi-billion dollar transition from a homegrown electronic health record system to a more commercial option through Epic. And so when they rolled out their new electronic health record, for the year just prior to that, we worked with Epic to create these new fields. And the thing is that Partners HealthCare was the second and largest academic health system in the country to have these new fields. Just a few months prior to our rollout, UC Davis was the first. So this is not something that is sort of widespread. Or at least, at the time, it wasn't. Since then, more institutions have started to roll out these structured fields.

So before that, how were they collecting the data?

Only in the clinical notes.

And that's just like typing in--

Free text.

Yeah.

And so if you only have that information in the clinical notes-- and you're an epidemiologist who's interested in studying health disparities. You're looking at cancer rates, anything of that nature. You're really hindered because you don't have an efficient way of identifying all of your LGBTQ patients, and then sort of studying their health outcomes. And so it really has tremendous downstream effects, not collecting these data in the way that we routinely collect other demographic data.

OK. So you're trying to improve on the quality of the data collection that primary care providers are doing around LGBTQ people so that you can more easily study that population?

Yeah. We can't even-- without these fields, we can't ask simple questions. Like, how are our LGBTQ patients faring in our health system?

Because you don't even know who they are.

Because we don't know who they are. And without having a baseline, we can't then compare that to any potential interventions in the future.

OK.

And so it may not be the sexiest topic, documentation, but the sort of utility of collecting these data is certainly tremendous. And also, just in terms of the one-on-one patient experience, having your primary care physician present a welcoming and compassionate environment, where you might feel safe to disclose this information-- I think that definitely signals things to patients, and makes them feel more welcome. At least, that's what we've found.

And you worked closely with your mentor, Dr. Zhou, on this project. Could you talk about how you began working together? And how you found each other?

Sure. So I've been a member of the Division of General Internal Medicine and Primary Care at the Brigham for a number of years. And I knew about Dr. Zhou's work. And she's just, does such fantastic research. But I was a little afraid to approach her because I always felt like-- I need to know more. I need to know more. I have to present a certain set of skills. Over the years, I've spent roughly seven years in that division. Over the years, I assumed those skills.

But it was at the nudging of someone else in our division who said, Harry, you need some more grant funding to fund your work. Dr. Zhou is an expert in exactly what you're interested in. And why don't you sit with her, send your CV, and see what happens. And when I first went to meet with her I was just so nervous and so junior. I'm still junior, but even more junior than. And so I approached her thinking I needed to have a bullet point list of all the things that I could offer her as a potential member of her lab. I was interested in doing anything that she might be doing. Of course, I had this interest in health disparities as well.

And to her credit, when I met with her, in addition to talking about her research, she noted on my CV, on my resume that there were a number of unfunded projects that I was interested in pursuing, but I just couldn't get the funding for-- in part because I was so junior. And she said, these are really interesting topics. They're really important topics. Why don't I help you carve out the time to pursue this research? I can pair you up with some of our postdocs and other students, and we can write some grants together. And you can work on the stuff that you really love because it's really interesting and important. And that's very rare. I mean, I've worked with a lot of wonderful people in my time, but I just was floored by that. And it's been a great experience for the last few years as a result. So I'm so thankful to her for that.

So Dr. Zhou, you work in artificial intelligence and natural language processing. Tell me what it was like when Harry first approached you, and how you started their mentoring relationship.

Yeah. So my lab is, as you said and as Harry said, my lab is doing a lot of work using natural language processing technology, which is a type of field of artificial intelligence that we look at. Particularly on the unstructured data as free text to data in the electronic health records. So we can mine information not presented the structured data field. Harry approached me with a different field, which is-- I don't have much experience, honestly. But we have the same interest in working on electronic health records, and the information technology.

I think using EHR data, we can better understand this population, and the care received, and how we use information technology to identify the important problems in this field, and promote the health care for this population, and understand what the patient's needs are. So it has been a great pleasure to working with Harry. Harry not have expertise in this specific field, but also has many other great skills that can help my team as well, like epidemiological methodologies. And also, like, temporal sense, we looking at longitudinal data-- those other things. Like, you know, has been a great addition to my team.

So Harry I just want to understand more about the study. Are you using artificial intelligence and natural language processing in this project as well?

Correct. Yes.

OK.

To get a little bit more into the nitty gritty, we had a few aims. The first was to describe and measure the extent to which SOGI, sexual orientation and gender identity, demographics were documented in the electronic health record using these new fields for the two year period following their introduction. And as well, to use natural language processing and machine learning to better understand that free text documentation. In addition to that, what we wanted to do was create these algorithms that basically use the new structured fields, all these data that are being collected over those last two years, to then identify specific phrases or terms within the clinical notes that might describe patients who are LGBTQ. And in doing so, we can increase the population that can be studied for quality improvement work and research.

And so in looking at the new structured fields over that two year period, from May 2015 to May 2017, what we found was that-- after reviewing 150,000 patients, or thereabouts, and over 400,000 visits. Among primary care patients over that two year period, only 21% of patients had a documented sexual orientation. And to give you a point of comparison for other demographics routinely collected in the health record, most of them are in single digits in terms of what patients have missing data.

And there were three fields, mind you, so not every field-- it's not an all or nothing game. Some patients have sexual orientation filled in, but not gender identity and sex assigned at birth. And so what we wanted to do was better-- one, just quantify that, because that's important. If you build it, they may not come. They may not use it. And so we want to make sure that if we're going to expend all this effort in creating these fields that people actually use it. And we need to better understand what goes into that decision making. And so what we did is we evaluated various predictors on whether or not a physician was collecting this information, or whether or not a patient was divulging this information.

And let me also point out that these questions are completely voluntary. So if a patient does not want to disclose this information, they have that opportunity. And if a physician doesn't want to ask this question, they also have that opportunity, as it currently stands. And so what that 80% missing represents is actually physicians who don't ask these questions. Within that 20%, a small minority of patients, when asked, did not disclose this information. So most people, when they're asked, are willing to tell you what their sexual orientation or gender identity is, as long as it's within the context of a clinical visit.

In addition to finding that only roughly 20% of patients had these fields completed. Even if that value was to say that they had declined to provide this information, there's still a lot of information there because that represents roughly 30,000 patients. That's a very large cohort, especially when you consider the lack of large scale research conducted to date, especially because of a lack of funding for this type of work. And so what we found is that in our population, roughly 4% do not identify as straight or heterosexual. And that sort of conforms to national estimates-- a little bit higher than 4%, and just under the 5% estimate that we have for the greater New England area.

What we also found was that, when asked, only about 1.5% of patients chose not to disclose. So a lot of patients are willing to provide these data. And previous studies also support this argument. So patients are willing to share this information. Other studies have demonstrated that they are able to answer these questions fairly easily. And so the barrier is really the onus of the health system to ask these questions and collect them in a uniform manner. And at the clinic level, this was really interesting. We saw a wide variety of practices. At the clinic level, we found that the percentage of patients documented ranged from as low as 7% to as high as 45%.

And documented, you mean in the SOGI demographics?

Yes. So I'm a physician. I have 100 patients. 7 of them would have their SOGI documented. And the other 93, there would be no information.

And do you-- I don't know. Maybe this is a question you haven't answered yet. But why is there that unevenness?

Yes.

Some people-- the physician marks the demographic info for somebody, but not another person.

So that's what we're trying to tease out at the moment. And we did various statistical analyses to try and identify independent predictors of whether or not a patient would have this information collected. And what we found is that the years of practice of a physician, essentially, physician age, or the number of years that they've been practicing medicine, did not matter. We thought perhaps that younger physicians might be more willing to collect this information. That wasn't the case.

We found that there was no difference in primary language. So a physician's ability to solicit information from a patient requires that they speak a common language. Right? Or there's some interpreter in the room, something like that. And we thought maybe that might play a role. But statistically speaking, adjusting for all these other factors, that wasn't the case.

But there were a number of other patient characteristics that were important, we found, for predicting whether or not a patient had those fields collected. Among them was the sex of their primary care physician. And we're talking about sexual orientation and gender identity. So when I talk about PCP sex, what I'm saying is the administrative sex. So this isn't necessarily how the physician identifies. This is simply information that's collected by the administration, and it's routinely available.

So looking at PCP sex, we found that patients with female doctors were 60 to 62% more likely to have their sexual orientation or gender identity collected. The older a patient was, the lower their probability of having SOGI documented. So for every decade older you are, you're 20 to 30% less likely to have SOGI documented.

So there are a number of possibilities there. Potentially, providers feel less eager to ask older patients these questions. Or older patients are less willing to divulge this information. But again, if they were less willing, we would have seen that in the declines, not just in missing data. We also saw differences in race and ethnicity, interestingly. So African-American patients and Hispanic, Latino, Latinx patients are much more likely to have SOGI demographics documented. This is very different from everything else we sort of see in terms of health inequity, health disparities in healthcare.

But interestingly, if you're an African-American patient, you were 16 to 24% more likely to have these fields documented as compared to white or Caucasian patients. And for the Latinx patients, they were 14 to 16% more likely. While in comparison, patients who are Asian or Pacific islander were actually less likely than their white or Caucasian counterparts to get these questions asked of them-- so 6 to 8% less relative to white or Caucasian patients.

Religion also played a role. We found that our Muslim patients were 22 to 23% less likely to get asked these questions. And patients not affiliated with a religion were also less likely, which was sort of an interesting thing, you know, dynamic there-- 8 to 9% less likely. And there are a number of factors that can play into that-- potentially that the provider not knowing the religious affiliation of a patient might be--

[INTERPOSING VOICES]

--afraid to--

Exactly. Less willing to. Also, income was a potential factor. Now for this, we don't, in our health system, routinely collect income of our patients. But we were able to link the median household income of a patient based on their zip code.

OK.

And found that patients who lived in wealthier areas were actually less likely to get these questions asked of them. So lots of different things we found based on demographic factors that sort of explain or predict whether or not a patient would have these data collected.

An important aspect of this grant was getting feedback from the community through a community engagement studio. Rodney, you were a part of the community engagement studio for this project. Could you tell us about your work, and your experience as a member of the community engagement studio?

I'm a public health practitioner. I work in health care. Worked for a number of years at Fenway Health, which is a community health center here in Boston that has focused its work on the LGBTQ population. Part of our work has been, in the early days, convincing people that collecting sexual orientation and gender identity information in the health record was an important thing to do. And then also, followed through with ways on how to do it. And how to train your staff. And get your policies and different structural things in order to make it to happen.

One of the things that was really cool about the process was, when I sat down with the other panelists, typically, I would have known everybody who was involved with the process. But I didn't. I knew probably two people out of-- I guess there were eight. Or there were a lot of-- so I think they really engaged a different group of people, which I find refreshing because it means that they just reach deep into the areas of expertise that it took to really review the different grants and stuff. So I enjoyed it, and getting to know new people. And it was really interesting to hear all these different projects that were being proposed. So it was really cool.

I think the most positive part about that is that you also end up-- so you get a review of research from a perspective. Like, well, how is this going to make my life better? Or how is this going to make the community's life better? Versus this is going to get somebody published. Or this is going to be up on a shelf, and nobody's going to pay attention to it. So I think that level of review was helpful in terms of making sure that it was really practical and beneficial.

As I can recall, we weren't deciders, but we were just input, offering input to the process. I do think there were some aha moments that I can recall, where people would say, hey, did you think about this? Or I'm experiencing this as that. And I think it, again, was helpful for people to hear that from outside of their own little bubbles of research or academia.

Could you talk a little bit about how this research project in particular addresses some of the health disparities that we see in the LGBTQ community?

I think the project was particularly interesting because it was looking at how people are gathering the information already. And was there a difference between the different patients who are being asked these questions? So I worked as part of a health care organization that no longer focuses on LGBT population, but the broader population. I feel like have a different lens on this today than I might have had when I worked in more of an LGBT focused environment.

What was interesting to me was that the techniques that the investigators were using are something that I think would be helpful just broadly in health care because it would be interesting to me to better understand what prevents providers from gathering the right information to provide the right care. So in this research study, it was really looking at SOGI data and the differences. But I'd be equally interested in things like weight, or height, or other pieces of health history that are also important because I think providers get so overwhelmed with having to gather information. How much of the lack of information around SOGI is about SOGI, or is about just our health care field in general?

With this information, you're getting towards a point where you can start predicting maybe which doctors, which practices are more or less likely to ask this information.

And depending on your clinic, you may be up to 10 times more likely to get asked these questions. So we had 16 clinics affiliated with one of our hospitals, or Partners HealthCare. And the range was essentially from 1 to 10 in terms of the 10 times more likely to have these data collected-- so huge differences. You can use machine learning, natural language processing, biostatistics, what have you. But we really also need to do qualitative examination of really what sort of goes into physicians asking these questions--

Yeah. Maybe you could talk more about the, sort of, like, how you use these technologies, like natural language processing. And maybe sort of limits of those approaches, and which would lead you to the qualitative idea.

Right. I mean, this area is sort of booming. They are new and interesting ways of leveraging AI to mine, for example, free text in the clinical notes. One of the things we did was we looked, we did something called word embeddings. I don't want to get too technical here. But essentially we used artificial neural networks to turn free text into numbers. OK? So the mathematical representation of a word. And we describe the relationship of those words to all of the words around them. OK?

And so the quintessential example of word embeddings is-- and then you can use math to sort of describe those relationships. So the quintessential example would be you can do you can take the mathematical representation of king, subtract man, and woman, and the result would be queen. Right? So all of these numbers, they're just represented in a document as numbers. You use math, subtract, divide, what have you, and these relationships hold.

So if you take a king, you change the gender, now it becomes queen. You can describe that mathematically. It's very fascinating, and a new-ish kind of area in the field. Past research has demonstrated that word embeddings are able to capture human implicit biases, or for example, gender stereotypes in a fashion parallel to the implicit association test.

You're saying that this technique, analyzing text using word embedding, can highlight implicit bias-- so somebody's stereotypes, prejudices, that kind of thing?

I think on a very basic level. Let's qualify that. But it can identify things-- so for example, past research has found that words like-- there's a famous paper out there were basically says that male gender is more closely associated with software engineer while housewife is more closely associated with female gender, to the same degree. And so there are various words out there in the world that are sort of more closely associated with gender that should be genderless. Obviously, housewife, wife, you would say that's female. But--

Engineer is--

Engineer is male. Art teacher is female.

Right.

And so using something akin to that, we can look at words that are more closely associated with LGBTQ people, or LGBTQ terms.

Can you give me an example what some of those are?

Yeah. So what we found is that, unlike for other demographic terms, race, ethnicity, things like that-- specifically for sexual orientation and gender identity, terms related to being LGBTQ, gay, lesbian, queer, things like that were more closely associated with terms like aggressive, chaotic, risky, discrimination, modifiable, reckless, disappointment, and abusive. And so there are a number of potential things going on here that we're just starting to delve into. But basically, it's possible that these words are more closely associated with patients because they're at greater risk for these things, for example, abuse. And some of the literature has borne that out in terms of interpersonal violence, things like that.

It's also possible that because these clinical notes are written by physicians, we're sort of delving into this realm of psycholinguistics, and it's really about perception. So a physician, a seasoned LGBTQ patient, maybe they don't have many of those types of patients. And so everything sort of goes within that lens of their own implicit biases when they're documenting things. Or some combination thereof. Or neither of those two things. But I think it does bear further exploration.

And I think regardless, it speaks to an issue of potential differences in terms of health outcome, health care access for our patients that we need to study more. Right now, what we've been able to do is establish that these terms, which are essentially synonyms, if you will, for these LGBTQ terms based on their mathematical representations-- what we can state is that within the electronic health record, these speak to potential issues, either health inequities or potential implicit biases. We can't be sure of that yet, but it definitely is a point, a jumping off point for further investigation.

OK.

Yeah. Also, I mentioned in the beginning that I have an interest in fairness accountability and transparency in artificial intelligence. And there's a real concern for bias in machine learning. And if we blindly apply these algorithms out there-- and necessarily, because of the way sexual orientation and gender identity are documented. If we're, even with the best of motives, using these techniques to identify a cohort of patients. If we don't keep in mind the potential for these biases, it may actually result in either perpetuation or exacerbation of further health inequity because they're sort of blindly applied. Because it can't be bias. It's a computer. Well, actually, research has demonstrated that there's a huge potential for bias if we don't examine it early on.

So in terms of its potential effect on care downstream, there may be some influence on the physician. We're talking about sort of like the clinical decision making aspect. But then also on the computational side as AI proliferates, there may be some issue there.

How do you hope that this project improves Health and where would you want to take this research from here?

Sure. So there are a few things. First of all, I think within our own health system, from a quality improvement standpoint, not necessarily a research standpoint, I hope that it spotlights a potential area for improvement within our health care system. We hope that people will use these fields, and providing these data will incentivize them to ask these of patients more often. Also, the sky's the limit in terms of using these information, these data. So there are a number of studies that have been sort of offshoots of this where we don't necessarily serve as the principal investigator or co-investigator, but they're using these same fields to further research. And like I said, the sky's the limit.

So hopefully it raises awareness that these fields exist and how to use them. We've paired with a number of projects to try and help them along that way. We've also paired, we've also worked with the Tegan and Sara Foundation to develop a Queer Health Hackathon using these data in a de-identified fashion to facilitate more interest, more research in this field.

And so I hope that there's both immediate improvement in care here. More people are talking about it. New areas of inquiry are inspired. For example, this whole word embedding piece. I also hope it inspires other people to choose this as a field. A lot of people, if they go into machine learning and natural language processing-- health disparities might not be the most natural next step or area of use using this information, these techniques, but I think there's a way to marry the two.

[MUSIC PLAYING]

Next time on Think Research.

I was brought up in Concord, New Hampshire across the street from the state hospital. And I used to see the psychiatric patients walking around on the grounds. And I just thought it was interesting that-- here's a group of people that other people don't act as if they exist. And what's going on? What's different about them? I mean, don't they have parents? Don't they have families? Why are they locked up?

Dr. Martha Shenton of Brigham and Women's Hospital explains how she came to study neuroimaging and schizophrenia. Thank you for listening. If you've enjoyed this podcast, please rate us on iTunes, and help us spread the word about the amazing research taking place across the Harvard community.

To learn more about the guests on this episode, visit our website Catalyst.Harvard.edu/ThinkResearch.