EP3_final.mp3

Adrian Hernandez: [00:00:01] 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.

Interviewer: [00:00:25] Today we're here with Greg Simon who just gave a fascinating presentation on bringing machine learning to the point of care to improve suicide prevention. Greg, it's great to have an opportunity to talk more with you about your experience. I wonder if you might like to tell us a little bit about where you think machine learning has the most opportunity to contribute.

Greg Simon: [00:00:48] Well I think the suicide prevention example is a good one because it's a rare event. So no individual provider will likely accumulate enough experience to really refine their own judgment. But if we can take the experience of thousands or tens of thousands of providers and add those together, we're likely to be able to come up with much more accurate predictions. The machine learning project that I described during the Grand Rounds really involves the information that healthcare providers already know and record. There's no deep mystery, there's nothing that health care providers can't see, there are no secrets. It's just that we're able to average across thousands or tens of thousands or sometimes hundreds of thousands of clinical judgements and make the best use of the combined wisdom of all the providers who've seen patients in these health systems.

Interviewer: [00:01:35] Now Greg you mentioned that the algorithm that you developed really focuses on data that the clinician is capturing, right, at hand there. Can you say something more about what other kinds of data you might bring to bear and what some of the considerations are.

Greg Simon: [00:01:51] We're starting with fairly simple data elements, the sort of discrete data elements that would be familiar to many people who know about traditional insurance claims or electronic health records. Things like ICD 9 and ICD 10 codes for diagnoses, procedure codes, medications that are dispensed. Already we know that there are things that providers know about that are not recorded in that discreet way. For instance providers may know really important things about people's lives, losses they've experienced, traumatic events they've experienced. Those could be very relevant to suicide risk but those don't tend to be recorded in a discreet way in the electronic health records. The providers may also record more subtle clinical judgment in the text of their notes. So someday it might be possible for us to mine the text of notes or even better to record those important life events in a systematic way because those would be really important for other health care providers to know. So I imagine we might be able to develop more accurate predictions if we had more accurate information about the things that actually happened between doctors and patients or therapists and patients in the consultation room.

Interviewer: [00:02:58] And is it possible that with that there'd be a tradeoff with respect to the transparency of the information that's going into the algorithm.

Greg Simon: [00:03:07] It is true that using the discrete data elements that we're capturing now it's pretty easy to describe what counts. We can say for instance that having a history of suicide attempt in the last year is a very powerful predictor of making another suicide attempt. And that makes sense to everyone. Or we can say that having a diagnosis of bipolar disorder is high risk, or even a diagnosis of a psychotic disorder is high risk. So that's simple and transparent and a provider can look at that and say that makes sense. If we were able to record life events in a systematic way it would certainly make sense to providers to say someone whose spouse has died or someone who's recently lost their job or who's experiencing financial pressures. Those people are at risk and providers already know that, that would make sense. It gets a little more complicated if we said we're turning machine learning loose on the words in text and we're finding particular words that matter, that might be a little harder for the human brain to comprehend.

Interviewer: [00:04:04] As you think about this algorithm that you've developed, or this machine learning algorithm that you've developed. What are your thoughts about how generalizable that might be to other clinical settings, other healthcare delivery systems and how would we go about thinking about how to evaluate that.

Greg Simon: [00:04:23] Well the model we developed was in several large integrated healthcare systems that serve fairly diverse populations. Many people across many different states, people with different kinds of health insurance, people with a wide variety of socioeconomic status levels and race and ethnicity. So we would hope that our findings are relatively generalizable. But it's certainly true that the way health care information is recorded can vary from place to place. So in our systems we find that more intense mental health treatment, certain diagnoses, certain procedures are certainly associated with higher risk. Of course we don't consider that to be a causal relationship it means that providers are accurately identifying people who are at risk and providing them more intensive treatment. If you getting data from a health care system where providers were not so attuned to that risk then the relationships might not be so strong. We have already posted most of our code on GitHub for people who had hoped to take our methods and apply them somewhere else. And what I would encourage someone to do is to first look at the basic distributions of the predictor variables. How common are these various things in our data. If the rates or the prevalence of these different predictors were wildly different that would certainly be a suggestion that these methods might not work so well. It would be possible for someone in another health care system, if they had accurate data about suicide attempts and suicide deaths, to simply take our model and see how well does it predict and how much slippage there is as you move from one place to the other. That's certainly something we hope to do, and have already entered into collaborations with some other different types of health care systems to start to test these models in more diverse settings.

Interviewer: [00:06:04] Can you tell us a little bit more about the intersection of this machine learning algorithm with the PHQ-9 which is already collected as part of care in your healthcare system.

Greg Simon: [00:06:17] I would describe the overall process as trying to develop a learning health care system to prevent suicide. Which means it's an iterative process. Our healthcare systems began to use the PHQ-9 standard self-report measure and we were able to prove that item 9 of the PHQ-9 was a strong predictor of subsequent suicide attempts and suicide death. So we began to use that and developed standard work processes in these healthcare systems to use that information to identify people at risk. But it was not as accurate as we had hoped. So we began to look at other types of healthcare data and ask how we could combine that with the self-report measures and that led to them machine learning models that I'm describing here. We're now starting to think about what are the other kinds of data that might be in the records. How could we use the data that are there but also how could we collaborate with the healthcare systems to improve the quality of data collected. We know that there are important things that happen to people in their daily lives that are related to risk of suicide attempts that aren't now recorded in any organized way. If those were recorded I'm sure our predictions would improve. So it's a back and forth or a sort of bi directional relationship. We collaborate with health care systems trying to answer questions using the data that are available now but also talk with them about how getting better data would serve the long term need of identifying people at risk.

Interviewer: [00:07:41] In this intervention, are patients are aware that a machine learning algorithm is being used to identify that they may be at risk of a suicide attempt?

Greg Simon: [00:07:53] What I would imagine would actually happen in an encounter would be that a provider alerted by this tool would conduct a more detailed assessment. Now the provider I suppose might explain why they were doing that but it's probably unnecessary. If you look at the risk models that we've developed, the things that those risk models pick are very obvious. We would be identifying people who have certain high risk diagnoses or people who have a history of suicide attempt or people who've had more severe mental health problems. So the idea of asking those people more specific and detailed questions about suicide risk would not be a surprise to anyone. It's not that there's any secret here or we're discovering something that would not be obvious to providers and patients already. We're just trying to steer people in the direction of not missing things and making sure that providers do their due diligence and don't overlook people who might be at risk.

Interviewer: [00:08:49] That makes a lot of sense. Our next podcast will be a moderator's edition podcast with Adrian Hernandez and Kevin Weinfurt and will be posted the week of November 13th.

Adrian Hernandez: [00:09:04] 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.