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. Thanks for joining.
Lesley Curtis:	00:27	Today we're here with Gregory Simon and Susan Shortreed, who will be reflecting on validating a computable phenotype: should results change a trial's pre-specified primary outcome? Greg and Susan, it's great to have you with us today. Greg, I wonder if we could begin by having you provide a little context about the trial in which this question arose.
Gregory SImon:	00:51	So our suicide prevention outreach trial was one of the first generation of pragmatic trials that are part of the healthcare system's research collaboratory. It's a trial of outreach programs to prevent suicide attempt. It's a very large trial. We've randomized about 19,000 people to either continue in usual care, or to be offered one of two outreach interventions. The outcome of interest is a self-harm diagnosis, likely suicide attempt that occurs in the electronic health record or comes in via insurance claims. Given the size of the trial, it certainly was not possible to interview people serially to ask them about self-harm events, or even to review detailed medical records from every person over 18 months. So we decided we had to rely on the diagnoses that were recorded. One of the wrinkles that turned up during the trial was we shifted from the ICD-9 diagnostic system to the ICD-10 diagnostic system shortly after the trial began. So even though we had done some work to define or to validate an outcome definition using ICD9, we had to do a hypothetical to project that forward into ICD-10 and hope that it would still apply.
Gregory SImon:	02:01	Of course, those data do come in in real time and we have the ability to monitor that. Susan, who's our lead biostatistician on the trial, has been doing some work on that and I'll turn it over to her and she can tell us we think we have learned a few things along the way.
Susan Shortreed:	02:20	Yeah, so we started looking at the event rate in the usual care to make sure we were on target with what we had intended and we were using our original definition that we had proposed in our protocol paper. But as Greg said, there was that transition from ICD-9 to ICD-10, where the individual codes that we were grouping up into our definition of suicide attempt changed. And so we actually looked at the number of events and the number

		of people who had events with each specific code and we divided those codes in kind of into three groupings. The first grouping is your more general, any injury or poisoning grouping. And then a second group would be any injury or poisoning with a self-harm diagnosis attached with it. And another third group is the more unusual group where we're unsure of how to incorporate this information into our primary outcome definition, is a group of codes that have an injury or poisoning event that have an "undetermined" attached to it. So it's not clearly an accidental or assault injury, nor is it clearly a self-harm injury.
Lesley Curtis:	03:32	Great, so have you learned anything along the way that's changed your mind about what you had initially planned to do?
Gregory SImon:	03:42	Well, how I would characterize where we are right now is we have not seen anything that tells us our original outcome definition was very far off-base. We haven't seen anything that tells us that we have a huge problem. But we certainly have learned that we might be able to do better. For instance, if we look at some of those diagnoses that we would have included, according to our original outcome specification, when you dig into the specific codes, some of them look to be unlikely. We would say ... It may sound silly, but one of the codes that occurred actually was a code for spider bite and we would say, "That's pretty unlikely to be a self-harm event or a suicide attempt." The question is, when we look at those relatively low frequency events which might contribute noise to our outcome, is that a big enough concern to now go back and change things? We might end up deciding it's not a big enough concern, we should stick with what we said. And certainly that's, in some ways, the simplest thing to do and the thing, to be honest, that raises the fewest concerns about us changing our outcome definition.
Gregory SImon:	04:49	But we may end up deciding that we now know that some of those events probably are not what we should be including. We could end up deciding to include or exclude some categories of events, just based on our suppositions. Using that example, we might say a wrist laceration or an overdose with psychotropic medication would have a higher probability than a spider bite to be a suicide attempt. Or we might end up deciding to actually review individual records in some of those gray area cases. Of course, we would have to be very careful if we decided to review records, that we were able to do that with a reviewer blinded to the intervention group, because it would be very problematic if somebody reviewing and classifying records had any knowledge of which group a participant was assigned to.

Lesley Curtis:	05:34	Thanks. You know, I wonder, stepping back from this trial and from the specific approach that you'll end up taking here, I wonder if there's some general lessons that you feel like you've learned throughout this process. And maybe I'll begin, Susan, with you, given the work that you've done with the data. Any takeaways for people who maybe be developing or about ready to begin a pragmatic clinical trial?
Susan Shortreed:	06:07	Yeah, I think there are some takeaways we've learned. One of the big takeaways with working with electronic healthcare records or data that were collected not for research purposes is it changes over time and it changes in ways that are outside of your control, often. And that includes both the actual data and then I would also say the data streams. So in order to get this information, we've done regular checks at the data to pull in information on codes. And there are times when I get a data set from a site and I immediately know that something's gone wrong, that we're not pulling in all of the codes. And these add an extra layer of complexity, even thinking about chart review and how things might change over time in defining your outcome and ensuring that you have all of the data you want for each individual in all three of ... In this ... Sorry, in our study, it's all three of the arms are control, usual care arm, and then our two intervention arms.
Susan Shortreed:	07:17	And I would also say, thinking about the data component and the statistical component of this idea of potentially chart reviewing outcomes in this setting. It's not just a binary setting that we often consider false positives and false negatives to. So as Greg said, we're looking at suicide attempt in the 18 months following randomization. And so me, as a statistician, that is a time-to-event analysis or survival analysis and we're looking, comparing the rates of time until first suicide attempt across the three different arms. And so when you consider maybe validating those self harm events and individuals having different number of events, the potential for planning a chart review can get quite complicated, especially when you're thinking about things, like Greg brought out, for blinding reviews and having to potentially scrub EHRs through a process so that all intervention material is cleaned out of the EHR so someone can be blinded. But if someone has an event that is then chart reviewed and hais said, "This is not a self harm event," but then they have a subsequent event two months later, then that should also be reviewed. And so it just adds a level of complexity that the statistician must be involved in in a really kind of deep level to make sure that you plan all of these complexities out from the beginning.

Lesley Curtis:	08:47	That's an important point, Susan, for sure. Greg, any takeaways that you would add to those that Susan identified?
Gregory SImon:	08:56	Well, I think the general point is that in pragmatic trials, where we're using what I call the data exhaustive healthcare, data that were generated by normal healthcare operations, as Susan said, they will change in ways beyond our control and we need to be continuously monitoring those data so that we might identify any potential problems. Some of those problems may be technical. It may be that something has happened in the data feed from this system to this system in this healthcare system and that's fixable. The data still exist. We say we've just missed some and we have to go back and find them. Some of them may have to do with changes in the informatics environment and as those of us who work in this area know, that how people record diagnoses can be influenced by all sorts of things: healthcare initiatives that might promote use of certain diagnoses over others, or even the financial incentives that might be attached to recording certain diagnoses. So it's important to be aware of those and when you see changes, to try to track those things down.
Gregory SImon:	09:57	Sometimes those things may be beyond your control. We use the specific, but time specific example of ICD-9 to ICD-10 transition, which was fairly dramatic. And that was one we were well-prepared for and sort of we knew we needed to monitor between October 2015 and the months before. But some of them may be, EHR or informatics changes may be less visible or apparent. So continuous monitoring is probably necessary to find out when something might change.
Lesley Curtis:	10:26	Great, well, thank you for making the time to talk with me today, Greg and Susan. It's really been a pleasure to hear more about your work. Please join us for our next podcast as we continue to highlight fascinating and informative changes in the research world.
Adrian H.:	10:49	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 PM Eastern time.