Institute for Research on Poverty
University of Wisconsin–Madison

Poverty Research and Policy Podcast

December 2016 podcast transcript

Featuring Tim Maloney, Auckland University of Technology (New Zealand), Head of the School of Economics and Co-Director of the Centre for Social Data Analytics

Hosted by Dave Chancellor

Using Predictive Analytics to Assess Risk for Child Maltreatment

Chancellor Hello and thanks for joining us for the December 2016 episode of the Poverty Research and Policy Podcast from the Institute for Research on Poverty at the University of Wisconsin Madison. I'm Dave Chancellor.

For this episode, we're going to be talking about how social service agencies, for example, can use administrative data as a tool to help predict risk among the people they serve.

To help us learn more about this, I talked to Tim Maloney who was visiting scholar at IRP this fall. Maloney heads the School of Economics at the Auckland University of Technology in New Zealand and co-directs the Center for Social Data Analytics with Rhema Vaithianathan there.

The Center for Social Data Analytics has already undertaken large projects both in New Zealand and in the United States involving predictive analytics and, to begin, I asked Maloney to tell us about the kind of work we're talking about here.

Maloney Predictive analytics is all about trying to use available administrative data to, as accurately as possible, predict these future outcomes of interest, whatever those things might be. So, in our case, for example, we're looking at historical relationships between these preexisting factors -- this is historical records, and say, the probability of substantiated maltreatment or the probability of being referred to social welfare, or having a child placed in foster care. So those are the sorts of things we care about, the sorts of things we want to predict, and we're just using whatever data are available in the most efficient manner possible to predict those outcomes of interest.

Chancellor Much of Maloney and his colleague's current work using predictive analytics stems from a project they did in 2013 for New Zealand's Ministry of Social Development or MSD.

Maloney So what we were asked to do was to use available administrative data in New Zealand to try to predict the probability of substantiated maltreatment of children. So the idea is that MSD said 'we've got a lot of data available and we don't know how to organize it to be predictive of the outcomes we care about and being able to monitor the maltreatment investigations and interventions that we undertake.' So they made the data available to us and we were able to come up with a predictive risk model that turned out to be quite predictive of potential maltreatment of children. And that's largely because of the quality of the linked data that we had available to us. So this linked data from the social welfare, from health, from judicial and criminal sources, from previous maltreatment for children and other allegations that were made… And we were able to use this in a systematic way to come up with a particularly effective predictive tool. We were hoping that this tool would actually be implemented and used by MSD in trying to investigate cases and to somehow rank the relative risk that individuals had in this area but up to this point it's met some resistance, both internally and externally to the ministry and it hasn't been implemented. But we think and hope that it will

Maloney, continued — be implemented at some point in the future.

Chancellor — Around the time they were finishing up this work in New Zealand with MSD, there was an RFP or request for proposal from Allegheny County in Pennsylvania, which is basically the Pittsburgh metro area, looking for a predictive risk model to help child welfare call screeners make decisions about the calls they were receiving.

Maloney — Allegheny County wanted something very similar to what the Ministry of Social Development wanted in New Zealand, which is to use this linked administrative data to come up with a predictive tool that they could use in call screening. So, the idea is that a call comes in to the hotline, alleging potential maltreatment and, obviously because of scarce resources they've got to make difficulty calls about whether or not particular calls get investigated and potential interventions occur as a result. They're actually obligated to look at all of the available data on these individuals to decide whether or not there is sufficient evidence to investigate. The problem is there is really a lot of data out there, there are lots of things that can be looked at and it's hard to summarize this, it's hard to obviously scrutinize all of the data that's available and the call screeners obviously don't know what's particularly predictive or what's not. So the idea is that we would use a statistical process to essentially combine all of this information and then give call screeners a single score which would allow them to know whether these particular cases are high risk or not.

Chancellor — One of the things that's allowed Allegheny County to explore this type of predictive analytics tool for their child welfare call screeners is the strength of the county's administrative data.

Maloney — Allegheny County is quite unique in the sense that they have a sort of integrated data management system a little bit like what we have in New Zealand. So it's not just a particular agency or department that's providing the data. We have data from a wide array of groups so this includes data on social welfare outcomes, past allegations of maltreatment. We have information about behavioral health issues, information from the hospital sector and so forth. It's really a large array of data that sort of gets fed into this predictive risk model.

Chancellor — Although this project was quite similar to their initial effort in New Zealand, Maloney cautions that you can't necessarily plug a new dataset into an existing model and set of techniques and expect it to be accurate.

Maloney — We've gone through a number of different estimation procedures to figure out what works best and I should say that one of the things that I think is kind of unique about our team is that we're quite agnostic as to what procedure or process we use in coming up with this risk tool. So we use the standard regression analysis, the econometric analysis that economists and a lot of other quantitative people would typically use but we're also considering nonstatistical techniques, machine learning, random force procedures, and, basically, we've said right from the outset that we'll use whatever tool works best, whatever is most predictive. So, rather than decide exactly what the methodology should be from the outset, we're exploring a range of methodologies that might yield much better results in the predictive process because in the end, it's all about being as accurate as you can in using the available data to predict these future outcomes.

Chancellor — The model Maloney and his colleagues developed for Allegheny County is now being taken a step beyond their New Zealand project; it is actually being used by the call screeners in fielding the day to day calls they receive. And then, the tool's effectiveness is being evaluated.

Maloney — We have an independent team that's going to do the evaluation so those of us who are responsible for developing this predictive risk tool basically step aside and let others come in and decide whether it's particularly useful or what potential issues might arise in the actual day to day use of this predictive risk tool, so I think it's an important part of the process that you really do need others who are sort of at arm's length from the actual development process to come in and do the evaluation work.

Chancellor — This evaluation can potentially help alleviate some of the concerns that have been identified with predictive analytics. Maloney notes that, although predictive analytics may offer a lot of promise, there's

Chancellor, continued

also discomfort around government agencies using administrative data for these kinds of purposes.

Maloney

It is a highly controversial area and we have had meetings with the general public, with interested parties to get their sense about what they are most concerned about in terms of using predictive risk analysis. And it comes in a variety of forms. One of the things I think that really bothers a lot of people is that it seems like we're potentially abetting or siding with the authorities that are providing limited funding to this particular area -- so the idea is that there have been cutbacks in funding and, as a result, maybe there's an attempt to use these predictive analytic tools to maybe more efficiently deal with the scarce resources that are available. We get the same criticism in university sectors, right? So you get a big cutback in your funding and you try to figure out ways of being more efficient, more cost effective, but I think those sorts of things are to some extent independent of the funding issue because in any funding regime, whether you have lots of money or very little money, you'd want to use things efficiently, you'd want to do things as best you can.

Chancellor

Maloney says there's also concern about the repurposing of previously collected administrative data for other purposes.

Maloney

So, people who are providing data aren't really informed about this potential alternative use for the information they're providing. But the important thing is that the data we're using is collected routinely as part of some other formal process. We're not talking about collecting new kinds of data, intervening more and trying to be more, I don't know, investigative, I guess, and providing different kinds of data. These are really data that already exist for some other purpose.

Chancellor

Beyond the data itself, there's also concern with the way the models are constructed, particularly regarding the possibility that they might reflect racial disparities.

Maloney

And that's certainly true when you're looking at allegations of maltreatment, child abuse, neglect, and so forth and it's one that we've taken on directly. One of the interesting issues here is do you include indicators of race in a predictive risk model? So you have lots of information about past outcomes, education, criminal actions, past allegations of abuse, but race is also in the model as well. What we've found is that once you condition on all of these other factors, once you have a complete history of things like alcohol abuse, treatment for mental health issues, allegations of past abuse and so forth, race turns out not to be all that important. So although there are racial disparities in a lot of these things, once you control for a wide array of other factors, race turns out to be not that important.

Chancellor

Maloney says there are ethical issues associated with this work, and not just those related to what happens with the data and with the related analysis once it is completed.

Maloney

So we have somebody from an ethical background who's already provided several reports on the work that we've done. And one of the things that's come up clearly in at least some of the reports is that there really is an ethical issue on the other end of the spectrum of, if you could do this with available data, not doing it is obviously a concern. So if you could identify somebody who's really high risk at the time of a call screening process and not intervening on that, not investigating on that outcome, that's potentially problematic. So that's an ethical issue. So you are sort of, there are ethical issues both in terms of what you do with the risk screening tools but also what you would do, and could do it, but didn't do it.

Chancellor

With that in mind, Maloney says that one of the interesting things about this work is to think about how a predictive risk tool performs relative to, say, existing ways of screening calls.

Maloney

So, there's already a process in place, I mean they have to make decisions currently without this predictive risk tool about who gets investigated and who doesn't, who gets screened in and who gets screened out. So, one of the things that we tried to do was kind of validate our approach relative to existing approaches by looking at objective outcomes like, for example, what are called Act 33 outcomes — deaths or near deaths from say, abuse that happens at some point in the future. The problem is it's a very limited way of validating because fortunately there are very few episodes of deaths or near deaths in our data set.

Maloney, continued

We're currently looking at hospital data with the hope that we can look at certain episodes that are reported as people are admitted to hospitals that might be indicative of maltreatment as an alternative way of validating our model because ultimately they like to know how well does your model stack up relative to the existing approaches used in call screening, and I've got to say that the preliminary work shows our predictive risk tools are substantially better, much more predictive of the things we care about.

Chancellor

The director of the Allegheny County Department of Human Services, Marc Cherna, said in an interview with PBS's NewsHour earlier this year that the potential for this tool is "just enormous" and that "this is the new frontier." Thanks to Tim Maloney for taking the time to talk about it with us.

This podcast was supported as part of a grant from the U.S. Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation but its contents should not be construed as representing the opinions or policy of that Office or any other agency of the Federal government or the Institute for Research on Poverty.

To catch new episodes of the Poverty Research and Policy Podcast, you can subscribe on iTunes, Stitcher, or your favorite podcast app. You can also find all of our past episodes on the Institute for Research on Poverty website. Thanks for listening.

[Intro and closing music is from "Test Drive" by Zapac, used under Creative Commons licensing].