

**Article:**

Ann M Moyer, Loralie J Langman, and John L Black, III.

A Novel Approach to Improve Accuracy of CYP2D6 Enzyme Activity and Drug Response Predictions

Clin Chem 2022;68: 616–618. <https://doi.org/10.1093/clinchem/hvab269>

Guest: Dr. Ann Moyer from the Mayo Clinic, in Rochester, MN.

Bob Barrett:

This is a podcast from *Clinical Chemistry*, sponsored by the Department of Laboratory Medicine at Boston Children's Hospital. I'm Bob Barrett.

Cytochrome P450 2D6 is a highly polymorphic enzyme that metabolizes many commonly prescribed medications. Because of that, it is frequently the subject of pharmacogenomic testing. In clinical practice, genetic biomarkers are used to categorize patients to predict cytochrome enzyme activity and adjust drug doses accordingly. However, this approach leaves a large part of variability in drug response unexplained.

A recent study from the Netherlands used genetic sequencing data and a continuous scale approach rather than categorical assignments to predict this enzyme's activity. A perspective article appearing in the May 2022 issue of *Clinical Chemistry* examined that work to improve phenotype predictions. The lead author of that paper is Dr. Ann Moyer, a Molecular Genetic Pathologist at the Mayo Clinic in Rochester, Minnesota, and Associate Professor of Laboratory Medicine and Pathology as well as Pharmacology. She is our guest in this podcast.

So, Dr. Moyer, what exactly is pharmacogenomics and what is the clinical significance of Cytochrome P450 2D6?

Ann Moyer:

Yeah, I think these are great questions. So, pharmacogenomics is the study of how genetic variation can impact the response to medications. So what we actually do clinically is we obtain a blood or saliva specimen from a patient, and then we extract the DNA, and then we perform a test to identify the genetic variants that that patient has, and then we can use that information to predict whether the patient might need a different dose or a different drug. And the goal is to avoid toxicity, but also to hopefully get the right dose of the right drug to give the patient the desired therapeutic response.

So CYP2D6 is a particular enzyme that's involved in drug metabolism and for those of us in pharmacogenomics, is one of our favorite genes to look at in part because it's really

complicated in that there's a lot of genetic variation or differences from person to person, but also it encodes an enzyme that's involved in the metabolism of almost a quarter of medications, and they range from anything from antidepressants to beta blockers to tamoxifen.

So, the drugs metabolized by CYP2D6 are involved in pretty much all areas of medicine. And so, it's a very interesting gene because of the technical challenges but it's also extremely clinically significant. So basically, we use this gene when we're applying pharmacogenomics in that we're doing genetic testing for CYP2D6 genetic variants, and then we use that to predict the activity of the enzyme and then apply that to medication and dose selection.

Bob Barrett: So doctor, let's talk about enzyme activity scores. What is that? How are they currently determined and how are they used in pharmacogenomics?

Ann Moyer: So, an activity score is basically a number that we use so that we can describe the relative activity of an allele, which is the form of the enzyme that's encoded by the genetic variance, as compared to the reference, or what we might call the wild-type. So, if you just took a typical person that had no genetic variance, their enzyme might have just what we would call a normal or full activity, and we would set that to a score of 1. And if somebody else had genetic variants that made it so that their enzyme didn't work at all, or at least one copy of their enzyme didn't work at all, we would set that activity at 0. And then if a patient has an allele with partial activity, we typically assign that to 0.5, or 50% activity.

So, for most genes, a patient should really have two copies of the gene, one from each parent. And so we take the score from each of those two alleles, or two copies of the gene, and we add them together. So, a patient that has two copies of the reference allele would have an activity score of 2 and that's what we would consider a normal metabolizer. And then we can use that activity score to translate it to the phenotype, and in pharmacogenomics what we mean by phenotype is you could be an ultra-rapid metabolizer with increased activity, you could be a normal metabolizer, or you could be an intermediate or a poor metabolizer with reduced or no enzyme activity.

So we like to use the activity score rather than just the phenotype because the activity score can be a little bit more granular than the phenotype. So for example, there could be a patient with an activity score of 1.5 or a patient with an activity score of 2, and in both cases, depending on the gene we were talking about, we might describe them as a normal metabolizer but in reality, that patient with the score of 1.5 might have a little bit decreased enzyme activity, and then

depending on the drug that they're being treated with, if it's metabolized by that enzyme and it has a narrow therapeutic index, so it means that it's hard to get the right dose to avoid toxicity and have optimal efficacy, maybe having that little bit reduced activity might actually end up being clinically significant. So, it just allows us to be a little bit more granular. So how we actually get to that number is, as I mentioned, basically where it's relative to a full or normal activity.

So you end up studying humans who either have a genetic variant and then observe their metabolic capability for certain substrates that are specific to that enzyme that you're interested in, or sometimes we do in vitro functional studies or enzyme activity assays, and then basically can experimentally look at the activity of the enzyme that you're studying with the genetic variance as compared to the normal variant that has no genetic variance within it and then that's how you set your activity score. And again, we usually right now are setting them at 0, 0.5, or 1 and then adding together the two alleles.

Bob Barrett: So, what is the difference between a categorical and a continuous activity score and why are categorical scores currently used?

Ann Moyer: So right now, what I just described are all really categorical enzyme activity scores. And so what I mean by that is basically we end up giving the patient a score of 0, 0.5, or 1 for each of their two alleles and then adding them together, but in reality, enzymes don't really behave that way. They can actually have an activity that falls anywhere on that spectrum from no activity at all to normal activity to even increased activity, but it's a little bit tricky to precisely measure that activity and to reproducibly measure it. So currently, for a genetic variant that results in reduced enzyme activity, we just say, "Oh, well, that's an activity score of 0.5." Even though in reality, what if that genetic variant made it so that you only had 30% of normal activity or maybe 60 or 70% of normal activity.

So presumably, instead of calling it .5, you could say, "Oh, okay. Well, it's 0.3 or 0.6 or 0.7." And if you did that, that would be more of a continuous activity score rather than basically rounding it to the nearest 0.5 which is what our categorical activity score is. And so specifically, for CYP2D6, it's particularly interesting because the field has worked really hard to come to consensus on how it handles CYP2D6, so it's incredibly important in pharmacogenomics. And so the professional societies and experts works together and they came up with a consensus for how you get from CYP2D6 genotype to phenotype. And in that consensus document, they also addressed the activity scores.

And I wasn't part of that particular group that decided to use the categorical activity score, but my suspicion is that probably some of the discussion centered around it being really difficult to precisely and reproducibly measure the activity so you could get to a really granular score. So for example, if you look at one study from one research lab, maybe they found that that variant of the enzyme had an activity of 62% of normal. So maybe you would want to assign a score of .62, but then a different group might have come along and studied the same thing and ended up finding that it had an activity of 73% of normal. So they would argue well it should be .73. So then what you do in that case?

So I think basically everybody can agree in a case like that: well, it's decreased. So maybe at this point, because there's difficulty in that precision and reproducibility with some of these studies to say, well .5 is pretty reasonable, or 50%, because we know it's reduced but it's a little hard to nail down exactly how reduced. However, there's still a little bit of disagreement as to whether a categorical activity score is really the best way to go or if maybe allowing other approaches with at least somewhat more granularity would be more appropriate.

So, there's actually a couple of alleles out there, actually specifically CYP2D6*10 has an activity score now of .25 because it was definitely lower than .5, but zero wasn't really appropriate either. So there's probably other alleles where this might apply too. And for example, in the recent article in *Science Translational Medicine* by the van der Lee group, where they're proposing a new approach to CYP2D6 prediction. They actually cited a study that showed that within the CYP2D6 phenotype category, so within intermediate metabolizers let's say, there's actually pretty significant variability in enzyme activity and they also cited the twin study that showed that although you can predict about 91% of CYP2D6 activity based on genetics or 91% of it rather is hereditary, the current system that we're using right now really only explains about 39% percent of the variability in CYP2D6 activity.

So, they're really suggesting that there's room for improvement and I think that's something that people in general are very interested in is can we do this even better? But for now the categorical scores are what we're currently using and they're working pretty well, but maybe we can do better.

Bob Barrett:

The original article that's the subject of the perspective used a new approach to determining the activity score. What exactly did they do?

Ann Moyer:

I thought they use a really cool approach. So, what they did is they actually used a neural network to predict CYP2D6 enzyme activity on a continuous scale, and they did it two different ways. So, they basically broke their model into two parts. So in one part, they used part of their model to generate activity scores just like what we might currently have, but they used a continuous scale and then they added together the activity scores for the alleles present, like we would normally, to get an overall activity score.

But then, they went a step further and they used the full version of their model where they actually didn't even call the individual activity scores for the two alleles but they just skipped right from this input data to predicting the final CYP2D6 activity. So, I thought that was really kind of an interesting approach that the first half of it is more similar to what we're doing now but maybe a different way to get at the activity score. And the second approach was, wow, you can use this neural network to do something entirely different to still predict CYP2D6 activity.

So to do this, they actually had three cohorts and I thought this was really interesting and it's kind of important to know how exactly it went about. But they had two of the cohorts were women with breast cancer and they were all treated with tamoxifen. One was a small cohort of men and women who are all taking venlafaxine, and all three cohorts were of European descent. So what they did is they took the first cohort as a training set for their model. So for the inputs for this model, they used the metabolic ratio of the parent drug to its metabolite and they measured that by mass spectrometry. And so basically, that's what they're using is their surrogate for a CYP2D6 activity. And then they also put in the genetic variants that were identified in each patient and then allowed the model to calculate out the activity scores.

So, basically what it was doing is the model was determining the impact of each genetic variant or combination of genetic variants in patients on CYP2D6 metabolism. And then they used their models predict the activity of the rest of their samples in their other cohorts based on the genetics. So the genetic information in that case was the input and then they compared the results of what their model was suggesting CYP2D6 activity should be to the actual measured CYP2D6 activity, or in this case, again, the metabolic ratios.

So after they did all of that, they could actually compare how their model performed compared to some of the conventional methods that we use today and what they found is that their model actually is able to predict CYP2D6 activity better than what we're currently doing and that's both if they use their full model where they're basically skipping over the whole activity score business, or the partial model where they

calculate the activity scores and then add them together. And then later, they went back to perform some functional studies to further validate the results of their model for several of their variants.

At the end of the day, they found that the model-based prediction was able to explain about 79% of the CYP2D6 activity, which was the best prediction in their study. And if they decided to just use the first part of the model and identify the continuous activity score that's predicted by that partial model and then add them together the way we do things today, that explains 73% of the variability in CYP2D6 activity. And then using their same data set when they used the traditional categorical activity score and phenotype, it only predicted about 67% and 54% of CYP2D6 activity. So, I thought it was kind of neat that both of their approaches could potentially improve CYP2D6 activity predictions.

Bob Barrett: Were there any limitations to their approach?

Ann Moyer: There were a little bit. So, I thought the approach itself was really neat. And I think it's a great proof of concept study. But one of the things that I thought was a little bit of a limitation was the specific cohorts that they had included. And so, I thought that could impact how we could use their results to our clinical practice today. So for example, tamoxifen is used to treat breast cancer and that, for the most part impacts women, although some men can also develop breast cancer.

So basically, in their original studies, two of those cohorts were women treated with tamoxifen. So, basically, it's mostly women and postmenopausal women, whereas their second cohort is again tamoxifen and pre or postmenopausal women. And then they had that other cohort of the patients treated with venlafaxine, which is used for depression. It was a relatively small cohort, but that one did include both men and women, but the limitation of that cohort is that all of these cohorts were of primarily patients of European descent. So, really, the limitation that I saw here is that the model was based on basically European women and a little bit of European men. And why this might be an issue is that specific genetic variants might be less common or more common depending on the patient's ancestral genetic background, and sex may also have an impact on drug metabolism.

So ideally, if you were going to use a model like this to identify the activity scores that maybe you could then pull out to use our typical methodology, but with these model-based continuous activity scores, it would really be better if you could build your model using a much more diverse cohort. But this was really meant to be a proof of concept study so I

don't think that was necessarily a limitation but definitely more of a future direction.

And in addition, although they did do a really good job with looking at the CYP2D6 activity for tamoxifen and venlafaxine, there are a lot of other drugs that get metabolized by CYP2D6 and it's possible that some genetic variants may impact different substrates to either a greater degree or to a lesser degree. And then also there's the possible influence of CYP2D6 inhibitors that that was really out of scope here too.

So, overall, I thought it was a really nice approach. But if we wanted to apply it clinically, I would definitely want to go back and make a much bigger cohort and then recalculate those activity scores.

Bob Barrett: Well, finally Dr. Moyer, how could this novel approach apply to a clinical laboratory setting?

Ann Moyer: Well, that's what I've been the most interested and thinking about ever since I first read this study. So, in the clinical laboratory, I think the easiest thing that we could do is we could potentially run a model like this again on a much more diverse, larger data set and then after you establish those continuous activity scores, maybe we could substitute them for our current categorical activity scores. And in that case, if we just did that part of the model, it seems like it was still quite predictive and the laboratory wouldn't have to figure out how on Earth to incorporate modeling into our day-to-day practice and we could still be providing results to individual patients in real time.

I think one of the challenges of that though, would be we just finally have the field in consensus about this is how we handle CYP2D6. And so, if we were going to suddenly change the activity scores, that would make a little bit challenging because labs would have to go back and make changes again to their tests. But if it ends up being better for the patients, maybe that's not a bad approach. But the other thing we'd have to do is the current model is really relying on using these specific large training set where the data for the level of the drug and the metabolite were used as those inputs.

So, I don't think it would have a great mechanism right now to predict the impact of a rare variant if we encountered one in clinical practice. So, I think it'd be neat if they could take this model and maybe add some additional features to it like perhaps some further work could add additional parameters, like, what's the amino acid substitution, and what are the properties of the amino acid that's normally at that position compared to the one that's introduced by the genetic variant? Or where in the enzyme is that substitution located? Is it in a critical domain? Or other variables, that would help better

predict the impact of a novel amino acid substitution, as well as the ones that were already part of the cohort and characterize.

So, I think there's still some work to do but this was a really exciting proof of concept study that raised some very interesting questions and considerations. And I think it's exciting because it means that maybe in the future, we'll be able to better predict enzyme activity and drug response phenotypes, and maybe some of that could be accomplished by utilizing modeling and some AI approaches like they did in this study.

Bob Barrett:

That was Dr. Ann Moyer, a Molecular Genetic Pathologist at the Mayo Clinic. She has been our guest in this podcast on a novel approach to improve the accuracy of Cytochrome P450 2D6 enzyme activity and drug response predictions. She is co-author of a perspective article describing that approach that appears in the May, 2022 issue of *Clinical Chemistry*. I'm Bob Barrett. Thanks for listening.