



Article:

M.J. Pencina. *Caution Is Needed in the Interpretation of Added Value of Biomarkers Analyzed in Matched Case Control Studies*. Clin Chem 2012;58:1176-8. <http://www.clinchem.org/content/58/8/1176.full>

Guest:

Dr. Michael Pencina is an Associate Professor in the Department of Biostatistics at Boston University and Director of Statistical Consulting at Harvard Clinical Research Institute

Bob Barrett:

This is the podcast from *Clinical Chemistry*. I am Bob Barrett. Biomarker research is at the forefront of the quest towards personalized medicine. It is hoped that the discovery of new biomarkers will aid in better understanding of disease, and in turn will lead to improved stratification of individuals at risk and to disease prevention.

However, in an editorial published in the August 2012 issue of the *Clinical Chemistry*, Dr. Michael Pencina writes that some degree of caution is needed in the interpretation of the value added by biomarkers. Dr. Pencina is an Associate Professor in the Department of Biostatistics at Boston University and Director of Statistical Consulting at Harvard Clinical Research Institute. He is our guest in this podcast.

Doctor, so far, the yield from the biomarker research in the cardiovascular arena has been somewhat limited. Do you think that the paper that you referenced in your editorial by Margaret Pepe and her colleagues about "Biases Introduced by Choosing Controls to Match Risk Factors" shed some light on potential causes?

Dr. Michael Pencina:

I think it does. I think it's an important paper and the premise of the paper illustrates the difference between the results obtained in matched case control studies versus prospective cohort studies and what Pepe is showing based on simulations and practical examples is that you can get a very promising apparent performance of a biomarker when using matched case control but then when you take that same biomarker to validate it in a cohort study, the performance becomes much worse and looking at standard metrics, the magnitude of deterioration is really big.

So translating into practical applications, we might get biomarkers that look promising or somewhat promising based on matched case control studies and many of the first line biomarker studies are and have to be matched case controlled and things look very good, and we get these potentially promising biomarkers and then in validation, they don't seem to offer much improvement to those prediction models.

Bob Barrett: Can the apparent performance of the marker be influenced by the age of the cohort under study?

Dr. Michael Pencina: Yes, that's one of the issues behind the difference and an issue that Pepe raises and illustrates. Basically what happens, age in cardiovascular primary prevention for example, is a very important risk factor which captures a lot of the information. And when we do a matched case control study and we match on age, we remove that importance of age, and what happens, markers look good because it usually is correlated with age, the impact of age is reduced and they look good. When we take them to cohort studies, where age is accurately controlled for, the performance deteriorates.

Moreover, what's also sometimes overlooked is the same marker can perform differently in two cohorts depending on their age distribution. There were two biomarker papers in The New England Journal of Medicine using similar biomarkers. One paper concluded that they improved risk prediction models and the other concluded that they don't. And the fundamental difference between these two papers were that one cohort had an age distribution as wide as forty years and the other as narrow as four years.

Bob Barrett: Well, should case control studies not be used for biomarker development? Is it feasible to just abandon them at the early stages?

Dr. Michael Pencina: Well, they are not ideal. The matched case control design has problems and these problems have been outlined by epidemiologists and I think there is generally agreement that if you can do a prospective cohort study, that's a much better design, but for feasibility issues, for the cost involved, I don't think we can abandon them entirely and in early stage of biomarker development, we have to use them.

But I think two things can be done. One is we might limit the number of variables on which we match, so avoid matching on age and the important

characteristics. So, do a case control study but not matched or it's little matched as possible, so that's one recommendation. The other recommendation is use the adjustment that Pepe proposed in her paper to account for the fact that matched case control or case control data is being used.

Farther more, what could be done and we are working on a theoretical model right now, that you can incorporate simulations in addition to characteristics of your marker which could come from case control studies. So you do a case control study, you take some partial results and you combine them with simulations based on real data coming, say, from Framingham Heart Study, which can give you a very accurate prediction of the potential of your marker, if it was used in a cohort study like Framingham.

Bob Barrett:

How should we evaluate the clinical potential of new markers?

Dr. Michael Pencina:

Well, that's a very good question. So, there are statistical methods and I think they belong to two main groups.

One is related more to the statistical measures and these measures are global in nature, they don't depend on thresholds or catalyst and the area under the ROC curve is one popular measure, the other one is discrimination slope. And so, I think, increase in the area under the curve or the increase in the discrimination slope, also called the Integrated Discrimination Improvement, are very good global measures and they have slightly different features, with the area under the curve focusing more on the relative risk and the discrimination slope and Integrated Discrimination Improvement focus more on the absolute risk.

Now, some people, especially many statisticians, like these measures because they are global and you don't have to introduce thresholds, which they find thresholds arbitrary.

There is a different view point or maybe it's a viewpoint in an earlier stage in biomarker evaluation, which looks at their clinical performance, which does depend on thresholds. And once you have thresholds introduced, it's really...you take sensitivity and specificity as the building blocks and then you can do different things with them.

So, one measure called net benefit coming from decision analysis provides the most appropriate weighting of importance between sensitivity and specificity or you can scale it - scale that benefit to get relative utility. So that's when you have one threshold. When you have two thresholds, then the Net Reclassification Improvement is, again, a measure that takes sensitivity and specificity at the different thresholds and combines it.

So these measures are kind of in between by going more into the clinical domain and the net benefit incorporates cost, which I think is very important but becomes very difficult because cost may not be known and may be changing. And then you can go to a formal full cost benefit analysis and build a big model but I think that has to wait after the first one or two steps have been cleared by a given biomarker.

Bob Barrett:

Are the statistical methods used to evaluate the impact of new markers overly conservative? Do they have any meaningful clinical interpretations?

Dr. Michael Pencina:

Well, that was the criticism for the area under the curve, saying that it's very difficult to increase it, once it reached a certain level and it's true, the mathematical properties of the area under the curve are such that once you get it to a level of say, 0.8, even a fairly strong marker is not able to increase it very much.

And it's a somewhat controversial issue and I think the best explanation to it is that there are two questions at hand. One is how much can we improve an existing model? And the other is are we dealing with a good promising marker? And it can be that in some applications, a marker that's fairly strong on its own will not improve the existing model very much, because the existing model is already very good. If we have a good model, it's natural that it will be harder to improve its performance. So the focus is a little bit different.

Now, if we go to a different setting, maybe a population with a more narrow age range, or maybe a population where we can obtain this marker more easily than some of the other variables included in the model, the importance of this promising marker might increase.

And so, it's important to see the difference between these two questions.

If the task is that discrimination separation focus is really on the model, are we doing better, then I think the AUC is still a pretty good metric. On the other hand, if we want to know more about the marker, then we might focus on different measures and the continuous Net Reclassification Improvement would be one measure that gives more focus and more importance to the marker itself, rather than the model to which it's being added.

Bob Barrett:

Well, finally, Dr. Pencina, do you think we really need new biomarkers in the cardiovascular arena?

Dr. Michael Pencina:

Well, I think there is always room to improve risk prediction models and the current performance of, say, Framingham function is around, this area under the curve is 0.76 or 0.8, with the maximum being 1, so there is still room to improve it. And people considered intermediate or low risk still suffer from events, so on that point, we need something to make the models better, and I think biomarkers are one avenue that can be very promising.

The other issue here is using the risk prediction models and the point in time in which they are being used. I think good models exist but they are underused by people, by physicians as well. There is still room to improve awareness of the risk of cardiovascular disease and the importance of prevention.

Also, the other issue is current models do depend on age and tend to point to older people being at high risk and that's true and it makes sense if we take the ten year perspective, but I think more effort should be made to evaluate the risk of cardiovascular disease in younger people, where there are more opportunities for prevention in lifestyle intervention rather than just treatment with medication.

Bob Barrett:

Dr. Michael Pencina is an Associate Professor in the Department of Biostatistics at Boston University and Director of Statistical Consulting at Harvard Clinical Research Institute. He has been our guest in this podcast from *Clinical Chemistry*.

I am Bob Barrett. Thanks for listening.