

Host: This is the podcast from the journal *Clinical Chemistry*. I am Bob Barrett. When prostate-specific antigen testing started to become widespread in the late 1980s, inconsistencies among results from different assays became apparent to some clinicians.

The December 2008 issue of the journal *Clinical Chemistry* published a paper from a group led by Dr. Jansen of the Netherlands, reporting use of an international WHO standard in the recalibration of a commonly used PSA assay. They found that the method of calibration affects the likelihood that a patient will undergo a biopsy and the probability that the procedure will reveal cancer. As a result, they propose a cut-off value as a clue for biopsy: a PSA concentration of three or four micrograms per liter.

The January 2009 issue of *Clinical Chemistry* contains a companion editorial on that study by Andrew Vickers and Hans Lilja. Dr. Hans Lilja is a Member and Attending Research Clinical Chemist in the Departments of Clinical Laboratory, Surgery, and Medicine, and Dr. Andrew Vickers is an Associate Attending Research Methodologist in the Department of Epidemiology and Biostatistics at Memorial Sloan-Kettering Cancer Center.

Together, they are questioning the contemporary or logic practice of using certain cut-points and are rethinking the relationship between biomarker measures and clinical decision making, perhaps even proposing the abandonment of the three or four micrograms per liter cut-point altogether.

Dr. Lilja, your editorial was written in response to a paper looking at the calibration of the PSA test. Why did the paper suggest whether a patient was advised to have a biopsy or not might depend on analytical testing procedures?

Dr. Hans Lilja: Well, the editorial was written in response to this particular paper written by the PSA group in Rotterdam and several papers dealing or addressing the same issue namely that when a novel standard, which has been or a calibrator for the PSA test, which has been endorsed by the WHO. When that is being used in laboratory testing that affects the way you interpret the cut-points because of the change in the values that are being reported. So there is an effect of about 20% difference in levels being measured with the same test, using or not using this calibrator.

Host: How are cut-points currently being used?

Dr. Hans Lilja: Well, they are standard in clinical laboratory testing, I would say. In that sense, they are ubiquitous, the standard format of our reporting assay results and interpreting the assay results. They are usually then constructed on the basis that you have a population on which you then delete either the top 5%, the bottom 5%, or you cut out the bottom 2.5%, and top 2.5%. It's just a statistical approach, which doesn't relate to the incidence of the disease or the severity or any outcomes related aspects.

Host: Okay. Well, Dr. Vickers, as a statistician, you, of course, understand the research behind cut-points. How are cut-points currently chosen? Why choose, say, four to determine abnormal amount of PSA rather than three?

Dr. Andrew Vickers: Great question, and the answer is we don't actually know. Dr. Lilja and I have actually spent a long time trying to work out where the cut-point of four actually came from, and we went back to the early history of the PSA test, and we wrote and we phoned, and then corresponded to some of the major figures around at the time, and all of them said, "I don't know," ask Smith, and Smith would say, "ask Jones," and Jones would say, "actually Brown knows," and Brown says, "well it all goes back to Smith."

This actually I think goes cross clinical chemistry. A little bit of the cut-points sort of found their way in to clinical practice go way, way, way back in time. Another example I know is some of the recommended daily amounts of Vitamin D come from the amount of a teaspoon of cod-liver oil.

So they are historical and not evidence-based and that's one of the reasons we think the cut-points are often so problematic.

Host: Well, cut-points are a part of routine care, of everyday care, why do you think they are so problematic?

Dr. Andrew Vickers: Well, a problem — in fact we don't know where many of the cut-points come from. So, we don't know their rationale, we don't know if they are good cut-points or bad. There's a whole slew of other cut-points, and we do know where they come from, and we think that the methods used to choose those cut-points were irrational.

[00:04:56]

Dr. Lilja has already mentioned the idea of a reference range. So, you take the population, you measure some marker or something in the blood, some substance in the blood, and then you say that if you are in the middle 95%, you are okay, and if you are in the top 2.5%, you are okay and you have some kind of a problem. If you are in the

bottom 2.5%, you also have some kind of problem and that would need further investigation.

Now better statistical concepts completely divorced from medical realities and medical clinical decision making. The example we give in our paper would be some — if there was some marker that was raised in children with sarcoma. Now if you measure this in little children, you would say, well about two million children in the US would end up being worthy of additional workup because there's about 100 million children in the US. In fact, thousands of cases, sarcoma in children are diagnosed each year.

There are a number of further problems with cut-points. One is that they are invariant to patient preference. Patients differ in how they consider the benefits and risks of various medical procedures and PSA is a very good point. One patient might come and say, "You know, I really don't think I can handle a biopsy, doc, you are sure it's really necessary?" And a doctor might say, "Well, you know, you are just above the cut-point of four, so maybe we won't bother." But how far above the cut-point should the doctor set a new cut-point of saying, "No, no, you really need to have this test now." Is it six, is it eight, is it ten, is it twelve. It's not really clear, it's not interpretable to a patient.

The final thing is, that of course with medicines becoming more sophisticated, we are getting a lot more information now from a lot more different sources, and a single cut-point can't include multiple pieces of information, and again, going back to prostate biopsies, we know not only that if you have higher PSA level do you have a higher risk of prostate cancer, and therefore, it might be worth having a biopsy. But we also know that the free-to-total PSA ratio isn't informative. So if you have a low free-to-total ratio that also may indicate cancer.

Now if you have two separate cut-points, it is difficult to explain to the patient whether they are a high risk or low risk. For example, you give a high PSA and also a PSA free-to-total ratio, does that mean you need a biopsy? It is not clear.

Host: So, what are the alternatives of cut-points?

Dr. Andrew Vickers: Let's think about what cut-points imply about biology. If we take, for example, PSA, there are two cut-points that are widely discussed. There is a cut-point of four which means you don't have a biopsy and a cut-point of ten which means you are really at high risk.

So the implication is that the way that makes you work is a series of steps, very, very low-risk, then suddenly higher

risk and then from four to ten, you suddenly hit ten, you are suddenly at very high risk.

Of course, nature and biology don't work like that. You have a smoother rise with risk, depending on your PSA level. So what Dr. Lilja and I recommended in our editorial is using very straightforward statistical methods all the time. Model the risk, work out what the risk of a specific end-point would be of your marker.

So, for example, suppose you had a lab report coming back, and say, this person has a PSA of six, which is out of range in high risk, and a free-to-total ratio of 17%, which is not out of range. There instead of that, we would combine those two pieces of information into a statistical model and tell the patient that they are at 25% risk, or a 15% risk, or a 40% risk of having prostate cancer and that percent risk would inform the clinician's decision whether to conduct a prostate biopsy.

Host: Wouldn't it make the clinical consultation really complicated to have to explain numerous possibilities to patients throughout the treatment, and would the doctors even understand?

Dr. Andrew Vickers: That's a great question. There is lots of evidence that people don't understand probabilities that well, and one of my favorite examples is someone who said, "You know I am either going to win the lottery or I am going to lose the lottery. So that's a 50/50 shot, right?"

I think, with all things in medicine, you can tailor the complexity of the patient-practitioner interaction depending on the situation. If you have some Wall Street type who deals with numbers all day and is quantitatively oriented, I can quite see the patient getting in to a sophisticated probabilistic discussion about whether the patient's risk warrant having a prostate biopsy.

There may be other patients who are less mathematically sophisticated and who are much happier doing, "Whatever you say, doc, I really don't want to be responsible for this decision."

[00:10:00]

In that case, the doctor could say, well, I could use some, kind of, national guideline. Right now, we say, biopsy above 4, don't; biopsy below 4. In place of that, we could say, just for example, we could biopsy above a risk of 20%, but not at a risk of 20% or below.

So I think there is enough flexibility in the medical system to allow for the use of probabilities and giving probabilities as part of the clinical chemistry report does not imply that they should necessarily be used explicitly during the clinical consultation.

Host: Okay. Well, Dr. Lilja back to you as someone who works in the clinical chemistry laboratory, do you think it would be feasible to have lab reports give those probabilities?

Dr. Hans Lilja: Yes, I certainly think it is feasible though, of course, we should acknowledge that given the practice we have today, where we don't have the systems implemented.

On the other hand, the technology is not so far reach, for example, of taking a spaceship to the moon. We have most of the necessary ingredients and the technology at hand, though, in need of implementation and proper validation of all the specific steps.

So, yes, surely, I could see this would be possible to cram around in a relatively short timeframe, given there would be more wide commitment and agreement on, but this is a necessary way to take it forward in the clinical laboratories.

Host: Lastly, Dr. Lilja, two studies that both have a large number of patients were published and reported on extensively in the lay press recently. *The New York Times* on March 18, had a headline that said, "*Prostate test found to save few lives.*"

Now, of course, you are familiar with these studies, how would your proposal to abandon set cut-points effect the interpretation of these reports?

Dr. Hans Lilja: Yes. Thank you for the question and certainly I am familiar with those studies, as I am a co-author of the European part of this study.

Let me first discuss the weaknesses and the strengths of the two studies and explain the viewpoint on that because I think it is important because they are very different in design and they are also, as was noted in the lay press, differences in conclusion from the two studies.

So, the negative outcome of the U.S. study, the so called PLCO Data is actually, in fact, largely explained by that the study is much smaller in size. It is somewhere between 70,000 and 80,000 men were included in the study compared to the plus 160,000 in the core age group of the European study. That's one explanation.

One is that there was a heavy pre-testing or screening before randomization in this population. There was low biopsy in response to PSA elevation in the study and also very important, there was a very high amount of contamination in the control arm of the group compared to the intent to screen the arm of the other group.

Lastly, there was a shorter follow-up. So, the medium follow-up in this study was only 7 years. So, all of these actually can then be used to explain, we believe that there actually should be absolutely no effects seen in this study.

Contrary then in the European study focusing then on a core age group of 55, 60, to 90 years at the randomization and was a medium on 9 years of follow up, and following an analysis which was the third interim analysis, actually two years ahead of the final conclusion of the study.

The study reached a conclusion that there was a significant and 20% reduction in mortality in the intention to screen compared to the controls. That is then actually 27% reduction in those who actually complied to screening, so when you adjust for the non-compliance, which was low. But still this analysis couldn't be done, then you see a larger effect which should be expected.

So that is the conclusion of the study that you then within these conditions of that amount of follow-up of 9 years whether it is large size, you actually see a significant reduction in mortality.

[00:15:01]

However, a side effect to that is then that you see a significant over-detection; you see a large number of — a very much larger number of cases diagnosed in the active screening arm compared to the control arm.

So, you will have an over-detection and a consequential risk of over-treatment in the controls, and there are numbers to this in terms of you need to treat close to 50 men to save one life. So, here is then the importance, I think, of discussions on how we asked with cut-points and with our biomarkers and that is to tell us that the current regimens on how we select these men is not optimal, and we need to change that.

One of the ways to change that is, of course, to put more emphasis as we have discussed in the cut-point on the long-term outcome and put the whole scenario into a risk-balance discussion where the risk of metastasis and death for the individual should be the main guidance on how you intervene on them.

If you then on contrary can identify a modality where your risk becomes very low, then you shouldn't intervene despite the fact that you may have, for example, a PSA level above a certain cut-point at a certain age in life. That's where this discussion comes in.

So, I think it's timely and appropriate in that setting.

Host:

Thank you, doctor. Dr. Hans Lilja is a member and Attending Research Clinical Chemist, and Dr. Andrew Vickers is an Associate Attending Research Methodologist at Memorial Sloan-Kettering Cancer Center, and our guests in this podcast from the *Clinical Chemistry*. I am Bob Barrett, thank you for listening.

Total Duration: 17 Minutes