Transcript | Threshold Video 3.27.2026

Today I'm going to discuss the importance of screening cut scores—also known as thresholds-–for decision making

To do so, I'm going to ask you to imagine a screening questionnaire that introduced in a previous talk. In this thought experiment, scores range between 0 and 9.

Score for children without problems, who I'll refer to as typically developing children are shown in blue. For these kids,

1. the average score is 4. Children with developmental behavioral problems are depicted in red,
2. and they have an average score of 5.5.

You can see the overlap even better if we flip the red distribution below the horizontal axis. And I'll also add a

1. cut score, which is also known as a threshold, which determines which scores
2. indicate a positive screen
3. Versus a negative screen

Among kids with developmental-behavioral problems, this distinguishes

1. true positives – from
2. false negatives . Likewise, among typically developing children, the threshold distinguishes between
3. True negatives –from
4. False positives –

and now we're ready to talk about the

1. tradeoffs are inherent in any clinical threshold—not just for developmental behavioral screeners – but for any medical test or any decision for that matter. As we raise the threshold,
2. Fewer children screen positive. This means that fewer children with problems screen positive, which reduces sensitivity. That's bad. However, it also means that fewer typically developing children screen positive, which raises specificity. That's good. It also means that more—or in this case most—of the kids who screen positive actually have problems. That is, PPV is high. That is also good. We see different tradeoffs when we
3. Lower the thresholds. More kids screen positive. This means that more children with problems screen positive, which increases sensitivity. That's good! However, it also means that more typically developing children screen positive, which reduces specificity. That's bad. It also means that fewer of the kids who screen positive actually have problems. That is, PPV is low. That is also bad. As I said, thresholds are a general feature of many decisions, not just medical tests and not just developmental behavioral screening. They also vary among different tests. Some screeners have very high thresholds—this may seem good at first because a high proportion of the children who screen positive turn out to have problems. But remember there are

tradeoffs—a high threshold typically implies lower sensitivity, which means the screener is missing kids you would like to detect. Conversely, some screeners—like the SWYC-- have lower thresholds. This may seem bad at first because they can yield a lot of false positives—but remember that the threshold was set where it was to ensure that sensitivity was comparatively high.

A standard approach to setting optimal thresholds relies on something called Youden's index, which basically says that the optimal threshold is the one that maximizes the sum of sensitivity and specificity. In the example, we've been discussing, the threshold is

1. in fact optimal according to Youden's Index. This means that
2. Sensitivity is 78%, so we will
3. miss 22% of children with problems. And it means that
4. Specificity is 78%, so
5. 22% of typically developing children with falsely screen positive. And it means that
6. Fewer than 50% of children who screen positive will actually have problems. What does it mean for this threshold to be optimal? Why are these particular tradeoffs the best? In my opinion, they seldom are, and I frankly think that the original article on Youden's index should be retracted. As a method, it is straightforward to apply, but it suggests that every point of sensitivity is as important as every point of specificity. Does this make sense? I think not. I think that Youden's index offers little more than a foolish consistency. But that's another story that I'd be happy to dig into in a later talk. For now, let's move on to a better way of thinking about thresholds.

Remember how some perspectives on evidence-based medicine emphasize judgment and values alongside evidence? Well, those perspectives are grounded in something called decision analysis. Here's a brief example of how decision analysis works. Rather than start with the evidence, decision analysis starts with a decision that needs to be made. In this case, we'll consider

1. A referral decision, but the same concepts can apply to any. For simplicity, here we consider only two choices: to refer or not to refer
2. Either way, there is a chance that the choice was correct or that it was not. Most decisions are made in the face of uncertainty, and that's what decision analysis is all about. IN this case, we're thinking of a referral as
3. correct if the child actually has a diagnosis and
4. incorrect if the child does not. Conversely, a decision not to refer
5. Is correct if the child does not have a diagnosis, whereas
6. it is incorrect if the child does. Finally, each possibility results in
7. A different outcome, in this case represented by estimates of quality of life. As your well aware, estimating quality of life is very much a question of values, and this is one reason why shared decision-making with patients is so important. Decision analysis and, by extension, health economics, recognizes this role for values. With some math, estimates of all of these values can be combined to calculate the
8. Expected utility of a referral and the expected utility of a decision not to refer. If one is definitely higher than the other, then the decision is clear. But if the two are equal, then we find ourselves at

Transcript | Threshold Video 3.27.2026

9. equipoise, at a point of indifference between two different choices. According to decision analysis, this is where the threshold is optimal. Now I totally understand that we often don't know the probabilities, and we can't simply rate quality of life on a numeric scale in a meaningful way. But nevertheless, decision analysis provides a useful framework for discussion. Here's how.

What all the math is saying is that if it truly makes sense to refer a child who scores above the threshold, this implies that the benefits of doing so

1. Outweigh the costs. In contrast, if it truly makes sense for a child [CLICK]
2. who scores below the threshold to NOT be referred. This implies that the costs outweigh the benefits [CLICK]
3. At some point in between, costs and benefits are in balance. This is the point of indifference, and by the logic of decision analysis, this is where the threshold should be. Let's consider two examples.

Let's start with cardiology. The USPSTF suggests that for those at risk of a cardiovascular event,

1. The benefits of taking statins far outweigh the risks and costs. So,
2. Where should the threshold be? Well, given their estimate of costs and benefits, the USPSTF says that the threshold should be
3. Quite low—even as low as
4. 8-10%. That is, if an individual has even an 8% risk of a cardiovascular event in the next 10 years, they should consider taking statins

Let's start with cardiology. The USPSTF suggests that for those at risk of a cardiovascular event,

1. The benefits of taking statins far outweigh the risks and costs. So,
2. Where should the threshold be? Well, given their estimate of costs and benefits, the USPSTF says that the threshold should be
3. Quite low—even as low as 8-10%. That is, if an individual has even an 8% risk of a cardiovascular event in the next 10 years, they should consider taking statins. Now you may agree or disagree—the USPSTF is quite clear that while these estimates are based on a host of evidence, they are also the product of expert judgment. And there is certainly room for shared decision-making with patients. But for now, let's contrast this with a different case: where should we set the threshold for an expensive and invasive surgery?

Even if the potential benefits are the same, the risks are considerably higher,

As are the costs

So the balance between the two is different. Does surgery make sense if the risk of a cardiovascular event is only 8-10%? Given the risks and costs, probably not

Transcript | Threshold Video 3.27.2026

A much higher threshold probably makes sense. So what have we learned by this example? Notice that I said nothing about how we are screening for cardiovascular risk—presumably we're using the same risk calculator in both examples. What changed is the treatment being considered, and more specifically, the risks, costs and benefits that are likely given that treatment. So an optimal threshold isn't really a property of a screening tool—it has much more to do with the net benefit of available options. That is, it depends on the context in which the screening occurs. In my opinion, this is

1. Key concept #6. Optimal thresholds requires weighing risks and costs with benefits. And while this process benefits from evidence, it clearly requires judgment and values. We can think of other examples that are specifically relevant to pediatrics. One such example is screening for risk of early onset sepsis among newborns The same logic applies, and the recommended threshold for administering antibiotics is again very low. And perhaps we can extend this same logic to clinical decision making for developmental behavioral problems

Imagine you are seeing a patient. You review the screening results.

1. Is it worthwhile to ask more questions? To probe a little further? To conduct more assessment? Decision analysis suggests that we should consider the costs and benefits of doing so. Now in this case, we have nothing like the level of evidence that is available in cardiology or for sepsis. But we could pose a few questions based on what we know. For example
2. Is a sensitivity of 78% adequate? That is, are we willing to miss
3. 22% of kids with problems without asking more questions?
4. And is a specificity of 78% acceptable? That is, are we willing to have conversations with 22% of families who have kids without problems even though it doesn't lead to any particular treatment or recommendation? Is the reassurance enough?
5. And are we ok with a PPV of almost 50%? That is, is it ok that further assessment will reveal problems in no more than 50% of cases? As a researcher, I can't answer these questions by myself. It depends on too many things that lie beyond the evidence at hand. For example, can pediatricians find the time, or is the burden too high? Will families in the community welcome the conversation? Are they likely to find it informative and even therapeutic, or will they find it stigmatizing? Setting optimal thresholds demands local perspectives

That said, I'll suggest that an assessment threshold should be comparatively low. Ideally, further questioning requires relatively little time, and the extra effort is worth it to ensure higher sensitivity—i.e., that as few children with problems are missed as possible. that said, exact thresholds will depend on one's judgment given local context.

But what about discussing treatment options. Would a clinician ever jump right to a conversation about treatment options based on screening results alone? Clearly not if the screening score is low. Here, it is useful to consider the

1. chance of having problems for children who score right at the threshold. In this case, it is vanishingly low—far too low to begin a discuss of treatment options. We would want the threshold to be much higher. How high? Again, that depends on local conditions

> and local perspectives, so I can't say for sure. But as an example, I'll point out that many evaluation clinics report that

they diagnose autism in about 60% of children who are referred. So perhaps this is a place to start. Compared to the previous threshold, the probability of diagnosis is much higher, as is the positive predictive value, which is now over 60%. That said, sensitivity is much lower—far more kids with problems are missed.

Put these two solutions together and we can see why two thresholds are often recommended. Two thresholds suggest three kinds of results—

1. Results indicating that symptoms and risk is high enough to justify immediate consideration of treatment
2. Another set of results suggesting that symptoms and risk are low enough to forgo further questions, instead reverting to standard surveillance
3. And a grey area in between where further questioning is indicated

To build on these general guidelines, remember our key concept from earlier. An optimal threshold represents a point where benefits can be expected to balance costs, and

1. Weighing costs & benefits requires judgment & values. Consider this question, what are the potential benefits of carefully assessing behavior and development for more patients? We've mentioned a clear benefit:
2. sensitivity

i.e., more children with actual problems will have the opportunity to benefit from further assessment. And let us not forget the therapeutic benefit that even a simple conversation can have. But now consider this question: what would you judge to be the potential costs? From a patient perspective,

1. more families will spend time on assessment, many of whom may not directly benefit.

and what about the clinician's perspective? Clearly,

1. assessment requires more time, and clinical time is a finite resource.

I argue that we too often ignore the value of clinicians' time, leading to longer and longer waitlists and increased burnout. Now these are only examples of possible costs and benefits that may apply to setting thresholds. Careful judgment is required to determine which outcomes are relevant. And even once outcomes have been identified, how should we balance the scales? Can we place a number on the value of families' and clinicians' time and compare that to the value of sensitivity? Some believe that utility can be measured with such precision, but I am not among the believers. Instead, I contend that the best we can do is use science as best we can to measure all relevant outcomes, and then work together to deliberate which threshold is optimal for a given clinic or health system. Use science to discern the truth--for example about the sensitivity, specificity, and PPV of the screener at different thresholds and perhaps even the time and burden on families and clinicians of various option. But even with the best available evidence, let's not forget that judgments about truth differ, so we'll need healthy deliberation to discern the good.

Transcript | Threshold Video 3.27.2026

I hope you this has been helpful. If you have more questions, check out our web site includes a range of information, including a technical Manual, scoring instructions, and translations. And check back for updates because we hope to add new information in the future. Thank you for listening