

**Article:**

Adith S Arun, Thomas J S Durant, Joe M El-Khoury, Harlan M Krumholz.

Reevaluating the Threshold for Low Total Testosterone.

Clin Chem 2025; 71(5): 609–11. <https://doi.org/10.1093/clinchem/hvaf025>

Guest: Adith Arun is a researcher at the Yale New Haven Hospital Center for Outcomes Research and Evaluation and a third year medical student at the Yale School of Medicine.

Bob Barrett:

This is a podcast from *Clinical Chemistry*, a production of the Association for Diagnostics & Laboratory Medicine. I'm Bob Barrett. If you've consumed popular media in any form recently, you've probably heard ads for testosterone replacement as the solution for men experiencing fatigue, low libido, or loss of muscle mass. In most cases, these symptoms can be attributed to a cause other than low testosterone, and hormone replacement is not warranted. In others, however, disruption in the hypothalamic-pituitary-gonadal axis results in true hypogonadism, and testosterone replacement may be appropriate for these individuals.

The clinical laboratory plays an essential role in establishing this diagnosis, as persistently low total testosterone is a key element of the diagnostic criteria. While this may seem like a straightforward concept, its implementation in clinical practice is complicated. First, different professional societies suggest different thresholds. Second, testosterone measurement methods are not interchangeable, and the percentage of men identified as having testosterone below a given threshold varies depending on the measurement method used.

With all of this uncertainty, how can we ensure appropriate care for all men undergoing evaluation for low testosterone? A new letter to the editor, appearing in the May 2025 issue of *Clinical Chemistry*, explores the recent history of testosterone measurement, summarizes the current landscape, and proposes reasons to re-evaluate the way we define low total testosterone.

In this podcast, we welcome the article's lead author. Adith Arun is a researcher at the Yale New Haven Hospital Center for Outcomes Research and Evaluation and a third-year medical student at the Yale School of Medicine. So, historically the cutoff for low testosterone has been 300 nanograms per deciliter. Could you walk us through why this threshold was chosen and how newer assay methods might prompt us to re-evaluate?

Adith Arun:

Yeah, that's a great question, Bob. So, I'll first start by talking about why the 300 nanograms per deciliter marker

was chosen. And so, there was some early work correlating essentially the first versions of these total testosterone measuring assays--they're called immunoassays--with the patient-reported symptom severity survey. So, basically, they just asked patients what their most troubling symptoms were and categorized them. And then they correlated the two of those together and were able to say that below a certain value the symptom threshold was quite high. And so, that became the basis of the standard.

And so, this was based on -- at that time, the most recent immunoassay work. But of course, now we're 25 years later and the methods have changed, so there have been newer immunoassay methods as well as mass spec methods as well. And so, people have worked to standardize them. And really the CDC has led that effort with their testing program that's accuracy based. And there have been a number of papers looking at the comparison between these two methods, but they just evolved so fast.

And I think today the mass spec method is now the standard across the community. And what we've found is that the current mass spec methods are quite concordant with their immunoassay counterparts. It's been a significant amount of work to make sure that that's the case. But there is variation, especially on the lower end of the spectrum between these assays, as well as the fact that these thresholds were set on versions of the immunoassay method that were made 25 years ago.

So, I think that's the historical basis for this 300 nanograms per deciliter. So, as with anything when you make a bunch of changes, and even though you are trying to make sure that they're standardized from one to the other, there will be some variation. I think this is just one of those times where there is this emphasis to re-evaluate it.

Bob Barrett: Well, when the cutoff isn't properly adjusted, how might that affect diagnosis? Could we be overestimating the number of men labeled with hypogonadism and potentially subjecting them to unnecessary treatment?

Adith Arun: Yeah, that's a good question. So, we actually -- just for some context here, we use the NHANES data set, which is this nationally representative sample of men and women, and we just focus on the men here. And we were just looking at the fraction of healthy males with total testosterone less than this 300 nanograms per deciliter.

And so, what we saw was that in 2004, it was about 12%. As a number of self-reported healthy males who had a measured total testosterone less than 300. And then in 2011, the next time that that was measured in the NHANES data set, it was

22%. And so, that's almost 100% increase from the 2004 metric. And that's interesting.

And we were able to see and dig into the data and show that it perfectly coincided with the migration of the assay measured, from the immunoassay to the mass spec method. So, that gave us a belief to believe that maybe this is maybe driven in part by the change in assay. And there have been other work that have shown that this total testosterone level decrease has happened in the context of controlling for different other variables.

And there's a different paper that recently shown the same set of data that total testosterone levels even decreased amongst males who are adolescent or young adults with a normal BMI. And they go on to state in this paper that additional factors could be contributing to this phenomenon, including things like diet changes or assay changes.

And so, we double clicked into the assay part of this equation. Coming back to this question of how it might affect diagnosis, so imagine that we have this assay change that that has happened and now individuals are now coming to get their testosterone tested for one reason or the other, and they see a low testosterone value. We're more likely to see a low testosterone value in someone who may just have a very normal testosterone level at baseline.

So, really this raises the concern that we could be overtreating things. And if this patient was to be prescribed testosterone replacement therapy, and this isn't a benign treatment. Testosterone replacement therapy has risks. There are thromboembolic risks. You can have some more blood clots as well as prostate hyperplasia that can eventually lead to something like benign prostatic hyperplasia as well.

And so, there are other consequences here as well. So, you can imagine that if you treat someone with testosterone replacement therapy, there's cost to the health system, there's cost to the patient, and there's probably delays in working up their true underlying cause.

Because a lot of the symptoms of hypogonadism that you can use in your history, along with a testosterone value that's low, are nonspecific. And it's things like having a low libido or fatigue or having a depressed mood. And that could be due to a number of causes, maybe that are medical or non-medical. And so, there are probably delays in workup that that could be happening. But really this is just a general trend that we've been able to observe.

Bob Barrett:

Well, guidelines are variable, with the American Urological Association suggesting 300 nanograms per deciliter, the

Endocrine Society recommending 264 nanograms per deciliter, and European bodies have their own cutoffs as well. How should physicians navigate these discrepancies in daily practice?

Adith Arun:

That's a great question, Bob. And I think ideally, we would have a consensus across all these different governing bodies. But of course, like you pointed out, there are different levels that are recommended by different groups.

And I think there are two broad strategies here that we can consider. One is we can bring awareness to the fact that there are these differences and that people should be aware of the changes in assays potentially affecting the cutoff levels. And so, once clinicians and patients are aware of this fact, you can have this appropriate level of cautiousness around interpreting these tests.

So, I think that's number one, really knowing when to use this binary on/off switch in terms of: above this level is good testosterone or a healthy testosterone and below this level is off, and sort of accepting that there's going to be a lot more uncertainty in this and that you have to -- like all clinicians do, incorporating the patient's history and symptoms into this, as well as their other exam findings.

And so, I think the second point, and the part of this strategy is that, I think it sets the stage perfectly to reconcile these discrepancies. One thing I just want to point out in this framework of reconciling these discrepancies is, if you go back to the NHANES set of data and you say, "Okay, what if I change the cutoff as the tests change?"

So, instead of it being 300 at 2011, which was based on the old assay, meaning that the 300 level, what if I change that cutoff to 264 and I recomputed the number of individuals or the percent of individuals with a low total testosterone below the 264 cutoff.

And what we do observe is that in 2013 onwards, we see 13%, 14%, and 11% of males who are self-reported healthy having a low total testosterone, which is in line with what we see in the immunoassay methods using the 300 cutoff in 2004. So, it goes 11%, 12% on the 300 numbers with the immunoassay, and then you switch over to the mass spec method and you say that the cutoff is now 264 and you see 13, 14, and 11% in the following cycles.

And that's just -- it's not proof of anything, it's an observation that we've made. It is an interesting observation and one that needs to be followed up on.

Bob Barrett: Well, we are in the era of personalized medicine. So, do you anticipate a movement away from a one-size-fits-all cutoff towards more patient-specific approaches?

Adith Arun: Yeah, in general, I think binary thresholds, like a certain level, maybe 300 or 264 are easy, but they're probably not the most informative. I think it might be nicer to say something like the probability of this patient's total testosterone being low given that the lab value we just measured is X is 60%. And I think that's probably more informative and more descriptive. But I think on the other hand, it runs the risk of not meaning much to clinicians. Like what does 58% mean versus 63% in terms of the probability?

So, I think in a personalized medicine world, I think we need to accept that binary thresholds are probably not the best way to operate, but they're good heuristics. And the right way to think about it is probably in terms of a more probabilistic sense, in terms of trying to assign probabilities like whether a patient's testosterone may or may not be low, given the data we just measured and the data that we have at hand, which is a Bayesian approach, but I think a lot of clinicians intuitively compute.

But I just wanted to harp on the fact that there is a lot of -- especially when you measure something like testosterone and really a lot of lab values, there's variation between patients, there's variation within a patient as the day progresses, there's the diurnal variation that we see in testosterone. So, I think appreciating that inherent variability in the data is key.

Ideally, we may establish a baseline for each person for whatever marker that we're interested in. We measure it from the time that they're really healthy and we track that over time, and we're able to establish a patient-specific baseline. But that's really just not a great use of resources and really not necessary, at least in this use case.

But I think at least we can compare them to the right comparison group. If this is a 45-year-old male, maybe we can compare them to other males of their age group that are matched accordingly, which probably give us the most informative understanding of where their testosterone levels are at.

Bob Barrett: Well, finally, looking ahead, what kind of research do you think is needed to ensure that guidelines and clinical practices align with the latest evidence on testosterone measurement?

Adith Arun: Yeah, that's a great question, Bob. And I think there are a few different cutoffs out there by different bodies, which you highlighted earlier in our conversation. And I think the first step is to try to harmonize these cutoffs. Because I think

asking physicians to navigate this on their own is not quite the right approach here. And I think getting a consensus definition is probably the best way to go ahead.

But as you pointed out, what do we need to do in order to reach a consensus definition? And I think we can use the understanding of the methods that we have today to inform what we do next. So, I think the fact that mass spec is what most labs use today, and it is considered the gold standard approach to measurement of this specific analyte. But certainly, there are some labs in different places across the world that use the immunoassay methods, like the newer generation immunoassay methods.

So, I think we need to harmonize these levels in the sense that we probably just need a single consensus cutoff point, given that we know, based on previous data that the current generation of mass spec methods and the current generation of immunoassay methods are very concordant to each other, as they have each individually gone through their own changes over the past 20 plus years. We now see that these two methods are quite similar to each other. And so, we just need to calibrate accordingly.

But I think to prove that these are aligned, we need to do that first in a patient-specific way where we collect data on patients who are probably healthy patients who volunteer this and measure it, both the immunoassay and the mass spec method, to establish the fact that these are actually very concordant. Then we just need to build a consensus definition of like, "Okay, is it 264, is it 300, or is it something else?" And I think that requires future work that is in the realm of lab medicine and clinical chemistry.

And so, I think that's an exciting new frontier to go to which is that, hey, we probably do need to pick a new cutoff or harmonize the cutoff and maybe pick. Maybe it's 264, maybe it's 300. And I think that is the next step to go.

Bob Barrett:

Adith Arun is a researcher at the Yale New Haven Hospital Center for Outcomes Research and Evaluation and a third-year medical student at the Yale School of Medicine in New Haven, Connecticut. He authored a letter to the editor in the May 2025 issue of *Clinical Chemistry* suggesting a re-evaluation of the threshold used to identify low total testosterone, and he's been our guest in this podcast on that topic. I'm Bob Barrett. Thanks for listening.