



Article:

Jillian Kodger, Joe M El-Khoury.

TSH Pediatric Reference Intervals: Lack of CALIPER Applicability to US-Based Populations.

Clin Chem 2026; 72(9): 1003–4. <https://doi.org/10.1093/clinchem/hvag075>

Guest: Dr. Joe El-Khoury is a Professor of Laboratory Medicine at Yale School of Medicine and Director of the Clinical Chemistry Laboratory at Yale-New Haven Hospital in New Haven, Connecticut.

Bob Barrett:

This is a podcast from *Clinical Chemistry*, a production of the Association for Diagnostics & Laboratory Medicine. I'm Bob Barrett. Much of laboratory medicine hinges on the designation of a test result as normal or abnormal. To provide value, a test result needs to be interpreted in the context of what's expected in healthy reference individuals. To provide this context, each lab test has at least one reference interval and potentially multiple if the analyte in question varies by age and sex.

Pediatric reference intervals are particularly challenging as concentrations of many analytes change over the course of childhood, requiring multiple different partitions for each analyte. Further, we generally don't collect samples from healthy kids. Recognizing this area of unmet need, the CALIPER database established pediatric reference intervals for a variety of analytes and assay platforms using samples from healthy Canadian children. The expectation is that these intervals would translate well to other geographically diverse patient populations, but a letter to the editor in the September 2026 issue of *Clinical Chemistry* points out that for at least one analyte, this was not the case. Why was the locally derived interval different, and what should laboratorians take away from this finding?

Today, we'll speak with that letter's senior author. Dr. Joe El-Khoury is a Professor of Laboratory Medicine at Yale School of Medicine and Director of the Clinical Chemistry Laboratory at Yale New Haven Hospital. He is board-certified by the American Board of Clinical Chemistry and a fellow at the ADLM Academy. So, Dr. El-Khoury, there is a lot to unpack here. Let's break this down for our audience who may not be familiar with the subject. Just what is CALIPER?

Joe El-Khoury:

Thank you, Bob. CALIPER is the Canadian Laboratory Initiative for Pediatric Reference Intervals. It is the largest prospective study to establish pediatric reference intervals in the world, and what's remarkable about what they've done is that they were able to recruit over 14,000 healthy children across the spectrum of age and used the samples from that population to establish reference intervals on five major

commercial vendors for automation platforms like Roche, Abbott, Siemens, Ortho Vitros, and Beckman. So, they were able to essentially develop pediatric reference intervals for a large group of chemistry tests on those platforms that were helpful to everybody in the world essentially to be able to use them because pediatric reference intervals are not as easily found in the literature on many of these platforms or provided by the manufacturers. So, this is a big deal for us in laboratory medicine.

Bob Barrett: Why is your finding that CALIPER's TSH ranges do not apply to a US-based population relevant?

Joe El-Khoury: Exactly. So, because, as I've mentioned, CALIPER is the largest study to date that has done this and they've done this remarkably well, many of us in the laboratory medicine community don't have resources to do what they did. So, what we end up doing often is referring to the literature and using reference intervals provided by those major studies for our own patient populations.

So, in this case, many laboratories, including in the US, have started implementing and using CALIPER's pediatric reference intervals. So, it's very relevant because many labs may be implementing the TS ranges that CALIPER is established up in Canada, and so our study essentially looked and said, "hey, wait a minute, these specific ranges may apply well to a Canadian population, but they did not apply to a US population when we've studied it in our study." So, it's a concern because, you know, as we go into the study, there's many things that can affect TSH that are weather-related or region-related, so it is important to make sure that that's accurate for your population, and if you're just taking CALIPER's ranges and implementing it for your own, it may not work as well, and you may have over-flagging of patients or under-flagging of patients. So that's why this is really relevant today.

Bob Barrett: Okay. So, can you briefly describe to us what you did in your study to help arrive at that conclusion?

Joe El-Khoury: Yes. So, with our fellow, Jillian Kodger, at the time, we basically looked at two years' worth of our outpatient data, so these are essentially pediatric patients in this case showing up to the hospital and having their blood drawn for a variety of reasons to assess their health that may not necessarily have an issue with their thyroid function. So, what we did is we looked at two years' worth of data, we were able to get thousands of patients' data points, and then we used a method that is called "refineR" that essentially looks at normal distribution of the data and excludes outliers when they're deriving the 95 percentile reference interval.

So, in less jargony terms, we used a big data set with cleaning tools that would allow us to reliably establish reference intervals, and we've essentially limited the impact of what could be people who could be sick presenting by only including one data point from each patient and essentially using this refineR technique that is known and established to manage these outliers that could otherwise affect our derivation.

Bob Barrett: And how does this conclusion fit with what is already known in the literature about TSH?

Joe El-Khoury: So, what we found essentially is that TSH in our population skewed lower than TSH in the Canadian population, and by quite a bit. So, for the age group 1 to 15, the difference was about 0.4. Ours was at the lower end, was about 0.7. CALIPER reported 1.1. So that is not unexpected, given that TSH is known to be higher in winter and colder climates. So, in Canada, it's expected that the range is going to be higher because they run a lot colder than we do in the US, and I will add, you know, I'm based in Connecticut, which is in the northeast. This difference may even be bigger going to the south of the US, like places in Florida, California, Southern California, at least, so where you're going to see these bigger effects. So, this is not surprising. It's just that many labs may not be thinking about reference intervals specifically for TSH and the seasonal impact when they're making these decisions.

But yes, the seasonal and climate changes with TSH are well published and known. So, this just fits that conclusion and says, yes, even for pediatric reference intervals, you need to take that into account. Otherwise, you're going to be over flagging.

Bob Barrett: It's remarkable that we've been testing TSH for so long, but its variation is not as widely appreciated. Could this problem be impacting other tests as well?

Joe El-Khoury: That was exactly what was surprising to me, Bob, entering this field and finding this. TSH is one that we've been obviously testing now for decades, 30 plus years, and looking at what the manufacturer provides us for reference intervals, even the one we use -- that study was done in like 1999, and we still use the reference intervals provided by them. And the concern there when looking at that study and making that decision for even the adult population, in my case, we discovered that they don't really account for these seasonal changes. So we have to basically have conversations with our endocrinology and add comments that it could go up in the winter. So if you're seeing flags in the winter, we recommend retesting two to three months later before taking action and prescribing, which in this case, they would typically

prescribe levothyroxine for the patient. So that is something that we've been able to address with TSH well, and now with this study, we did it for pediatric patients. Other tests that are affected by season, which there aren't a lot that are routinely done, but could also see this problem. Like if I was to think of like tests that would change by season would be like vitamin D obviously because it's higher due to when you have more sun exposure, so you'll typically see higher vitamin D levels in the summer and lower in the winter.

Similarly, cholesterol, A_{1c}, and glucose also change where you have them being higher in the winter, lower in the summer, and that's obviously because you have typically lighter diets in the summer and you're more active, so those changes are expected. So, one tip for people who are having higher than I expected cholesterol or glucose is just get retested in the summer and it'll be lower. No, joking aside, the reason that you don't see as much of an issue with those is because we have clear guidelines, and those are standardized tests that establish what is acceptable. Whereas with TSH, there isn't. In fact, reference intervals for TSH vary by method, and they don't account for the seasonality. So TSH is the biggest problem we have, but absolutely, tests that are affected by season will also see this change.

Bob Barrett: Well, finally, Dr. El-Khoury, given what you report in your study and the difficulty of developing pediatric reference intervals, what do you recommend laboratories do when implementing pediatric reference intervals?

Joe El-Khoury: Yeah, so I sympathize with many labs who don't have the resources, including us to do like prospective studies with healthy children. Guidelines recommend local verification of reference intervals which is what we did in this case for many of the tests on the comprehensive metabolic panel that Dr. Gajjar and I published on separately, and then here with thyroid tests, so we looked at TSH and FT₄. But it's also difficult to do verification as recommended by the guidelines to get 20 healthy individuals, because you're not going to easily find 20 healthy children in each age group that you have at reference interval to donate blood for a study, right?

It's not something that -- with adults, it's easier, with children, it's very difficult. So, my recommendation to labs is before implementing the reference intervals from the literature like CALIPER's, you could use your outpatient data which you already have in your laboratory to study, using tools like refineR and compare and make sure that the CALIPER's intervals match your data in your own outpatient hospital.

So, in my experience, this has been the most cost effective because we already have that data. It's not costing us any

additional dollars to run the agents or tests or even go out there and do essentially clinical trials by recruiting patients to donate blood for these studies. It costs us nothing except obviously someone who has experience using refineR and doing the data analysis to generate data that shows that this works for you. So, my recommendation and lesson learned here is before implementing these reference intervals from the literature, verify them that they work for your population using these outpatient data approaches. They work well, especially in tests that are commonly ordered.

Obviously, I will caution and say they will not work well for tests where you have more diseased patients presenting and that could impact those tests. Generally, for chemistry, at least, we get a lot more patients with normal results for a comprehensive metabolic panel, thyroid panel than not, so those approaches are very useful. And, so, lesson learned here also for TSH specifically, since that changes by location due to climate and latitude, I would strongly urge cities in the US, hospitals in the US, in different cities, to look at that carefully and not blindly implement CALIPER's because it's way higher for a colder climate and definitely would be skewed high for cities in the US and even in other places of the world equally important where the climates may be different. So, I would recommend essentially locally verifying those ranges using your outpatient data with a tool like refineR.

Bob Barrett:

That was Dr. Joe El-Khoury from the Yale School of Medicine in New Haven, Connecticut. He wrote a letter to the editor in the September 2026 issue of *Clinical Chemistry* describing discordance between locally derived and database reference intervals. He's been our guest in this podcast on that topic. I'm Bob Barrett. Thanks for listening.