Thank you, Dr. Chalasani. The title for this talk was not my idea. It was Dr. Senior's idea. I am not sure what "Down with the Tower of Babel" means, but I am going to try to continue on with that, anyhow.
So, if we could have the next slide, please? Unfortunately, I have no relevant conflicts to disclose. (Laughter.)
Why do ALT results differ between different laboratories? We heard some of the issues about this from Dr. Prati. There may be differences in the approaches that are used for establishing reference values. Additionally, there may be differences in the methods used, and there may be differences between manufacturers that contribute to differences in results.
So, as we heard from Dr. Karmen, the method for measuring AST and ALT has not really changed very much since the assay by Dr. Karmen was introduced over sixty years ago. But, as they differ, you may remember learning, if you took any biochemistry, that the activity of enzymes is influenced by a number of things: pH, ionic strength, temperature, concentration of the reagents that are present, the buffer that is used. In addition, for both alanine and aspartate aminotransferase, pyridoxal 5′ phosphate, which is a vitamin, is needed as a cofactor. But many manufacturers do not include optimal amounts of this in their reagents. There are reasons for why they don't do this, because it reduces the values for which you don't have to dilute the sample and, additionally, it affects the stability of the reagents. But, a lot of laboratories don't use it in their assays.

**HOW IS ALT MEASURED?**

- All laboratories use a form of the assay developed by Karmen, which measures enzymatic activity
- Assays may differ in a variety of parameters that can affect this activity, such as pH, buffer, concentration of reagents
- A necessary co-factor for activity is pyridoxal-5’-phoshpate (P-5’-P), which is not added by many manufacturers
All laboratories in the United States are regulated by the Food and Drug Administration or by the Centers for Medicare and Medicaid Services. They are required to test unknowns in a procedure that is called proficiency testing (PT). The U.S. law sets a requirement of reproducibility. In other words, if the true value is supposed to be a certain amount, there is a limit for how far an individual laboratory can be off and still considered passable. That limit, set by law, is plus or minus 20 percent. So, that is a pretty broad range that a laboratory can be within and still be considered acceptable.

Moreover, laboratories are usually compared against all other laboratories that are using the same instrument. So, as long as you are within plus or minus 20 percent for what everybody gets as the average with that method, that particular manufacturer's assay, then you are considered acceptable. So, there is a pretty low bar set for passing for ALT.

HOW COMPARABLE ARE ALT METHODS?

• In US, laboratories are required to measure unknowns, and report results, termed proficiency testing (PT).
• US law sets degree of agreement needed to pass PT, which labs generally consider acceptable performance
• Acceptable results for ALT are ± 20% of the average of all labs using the same method
Now in the United States the major provider for proficiency testing materials is the College of American Pathologists (CAP). In their most recent data, ALT was performed by about 5200 different laboratories using instruments from one of five major manufacturers. There are some more minority groups, but these are the five main manufacturers.

Looking at a couple of samples that have mildly elevated ALT, the average that was reported by these different instrument manufacturers for one of the samples ranged from 59 to 81, and on the other from 76 to 99. So, these are the average values. And remember, labs can be plus or minus 20 percent around that average and still be considered acceptable. So, there is, again, a pretty broad range of values which you might encounter in testing the same sample.
Now results from the same manufacturer are generally more comparable, but different manufacturers have different platforms and different kits they can have. So, for example, with most of them, the results were more agreeable and in the range of 70 to 90 with the pyridoxal 5' phosphate was in there. Most of the results on average were within about two or three. And they were similar even at higher ALT results as well.

So, there is a lot of disparity among the different methods for ALT using these unknown samples. Now I will say, as a laboratorian, that the samples that are tested have to be stabilized in some way. And so, sometimes they don't perform exactly the same way as real patient samples would. There are things that can happen with them, but this is the best data that we have on how well things agree.
Dr. Chalasani published a study a number of years ago in Hepatology on looking at what laboratories used as their reference values and found that they also differed markedly from one laboratory to another.

So, how do laboratories generally establish reference intervals? Well, you heard from Dr. Prati on this that the minimum requirement that laboratories have to do is validate the reference interval that they are using, the, quote, "normal range".
For a laboratory to establish its own reference interval, it is often very difficult. So, most of them just try to validate what the manufacturer suggests. That leads to a further variability in what the upper limits of normal are from one laboratory to another. So, this is a problem that occurs still.

REFERENCE INTERVALS

- If manufacturer has not considered the factors discussed by Dr. Prati, then proposed reference interval may be wider than what would be reasonable for “healthy” values
- Some laboratories select their own sample, which can lead to different results as well
- The combination of different methods and different labs making reference intervals makes for lack of uniformity
Now what are some approaches that could be taken in addition to that? We have heard some theoretical ways of doing this from Dr. Prati, but I would like to share with you some experiences that have been done with other tests to try to reach some sort of agreement. So, let me take you through a few of these.

**WHAT CAN BE DONE?**

- There are informative lessons from other tests that could be applied to ALT to make both actual results and their interpretation more uniform
- Will require cooperation between clinical and laboratory organizations to make this happen
The first one of these is cholesterol. So, back in the early 1980s when I was a resident and some people that are my age or older may remember this that our, quote "normal values" for cholesterol for people in their sixties, such as me, would go up to about 340. Anybody who is younger than that and has heard anything about cholesterol knows that we don't think that is normal anymore.

Well, back in the 1980s, the Heart, Lung, and Blood Institute established a National Cholesterol Education Program. They wanted to try to reduce the incidence of cardiac events and make cholesterol results more comparable among laboratories.
So, part of what was done here was to have a laboratory effort to improve reproducibility of cholesterol values among different laboratories. And risk values were defined based on cholesterol and LDL levels that in prospective studies, which you heard about from Dr. Prati, had been used to establish where the cutoffs would be to show an increasing risk of heart attacks occurring.

To give you an idea how well the reproducibility has improved, in the most recent proficiency testing survey using samples that had average values about 190, the averages among those five different manufactures ranged between 183 and 196, so a lot less variability than with ALT, but most of them had averages between 190 and 194. So much improved reproducibility among laboratories with cholesterol than used to be the case 20 or 30 years ago.
Another example is hemoglobin A1c. In the 1990s the Diabetes Control and Complications Trial established that A1c values were important in predicting microvascular complications in patients with diabetes. However, A1c was never used as a diagnostic criterion and it was felt to be problematic even for monitoring patients because, like the situation that exists with ALT currently, A1c values were not in agreement among different laboratories.
And so, as a result, there was a lot of effort to try to improve this. There was the program called the National Glycohemoglobin Standardization Program that was set up to work with manufacturers to improve repeatability of A1c values using different methods.
Currently, to be certified by the NGSP, manufacturers have to have results that are within 6 percent, not percentage points of their result, but within 6 percent of the true value for the given sample. So, for example, if A1c was actually 7 percent, the manufacturer had to report A1c's between 6.6 and 7.4 percent, and they have been gradually tightening this range of what is acceptable.

So, based on this improved agreement, in 2010 the American Diabetes Association adopted A1c as a diagnostic criterion for diagnosing diabetes. What was used as the threshold for diagnosing diabetes was the value based on prospective studies where the risk of developing microvascular complications, in this case retinopathy, began to significantly increase. So, again, a clinically-defined decision limit for A1c was used after laboratory results were standardized.
Now this led The Endocrine Society to look at other laboratory tests and say, can we improve the reproducibility of these results among different laboratories? And so, they developed a program called PATH, which is Partnership for Accurate Testing of Hormones (PATH).
To date, they have worked with a number of hormones, including testosterone and estradiol and vitamin D, and are close to having results and good agreement among different laboratories.

**PATH**

- To date, manufacturers and large laboratories have worked to develop a program to make results agree more closely for a number of hormones (testosterone, estradiol, vitamin D)
- The plan is to expand the effort to other hormones over time
So, in summary, ALT results can currently differ significantly, although the difference is less in laboratories that use the recommended method and add ideal amounts of pyridoxal 5’ phosphate. Most laboratories, however, don't use that.

There are alternative approaches to improve the repeatability of testing, but this has really required intervention from clinicians saying, "This is not acceptable to us. We need to work together between the laboratory and the clinicians to develop targets for what is appropriate for deciding whether disease is present or not and, also, for laboratories to work with manufacturers to improve the repeatability of the results among different laboratories and different methods as well.
So, really, it is going to require a cooperative effort where the laboratory associations have to be prodded by the clinicians. This means, for those who are clinically-inclined in the audience, you need to get involved. You need to make your voices heard and say this is not acceptable. We need to work together in an appropriate organization to do this. It might be AASLD.

But everybody needs to work together to improve the repeatability of ALT values and really make it possible for us to set a cutoff where we can recognize drug-induced liver injury. So, really, we need cooperation and a motivating factor to get people to work to improve the repeatability of these results.

**SUMMARY**

- This approach requires support of professional societies of clinicians AND laboratorians
- Together can address both health implications and laboratory procedures needed to assure comparable results between labs