FOOD AND DRUG ADMINISTRATION

CENTER FOR DRUG EVALUATION AND RESEARCH

OFFICE OF SURVEILLANCE AND EPIDEMIOLOGY

+ + + + +

DRUG-INDUCED LIVER INJURY CONFERENCE XV

The Importance of Getting It Right
How to Measure & Interpret DILI Information, Make Correct Diagnoses

+ + + + +

WEDNESDAY
MARCH 18, 2015

+ + + + +

The Conference convened in the University of Maryland Marriott Conference Center, Chesapeake Ballroom, 3501 University Boulevard East, Hyattsville, Maryland, at 7:55 a.m., John Senior, Paul Watkins, Mark Avigan, and Lana Pauls, Organizers, presiding.

PRESENT

JOHN SENIOR, Organizer; Speaker, Session IIA2
PAUL WATKINS, Organizer; Moderator, LSRC
MARK AVIGAN, Organizer
LANA PAULS, Organizer; Speaker, Session IIA1
SOLOMON IYASU, Introductory Speaker
ROBERT DUFOUR, Moderator, Session I, Speaker IA3
NAGA CHALASANI, Moderator, Session I, Speaker IB3
ROBERT TEMPLE, Moderator, Session II, Speaker IIB2
ARTHUR KARMEN, Speaker, Session IA1
DANIELE PRATI, Speaker, Session IA2
NIRA POLLOCK, Speaker, Session IA4
LEONARD SEEFF, Speaker, Session IB1
TED GUO, Speaker, Session IB2
PATRICK KIRBY, Speaker, Session IB4
ALICE CHEN, Speaker, Session IIA3
KAREN HICKS, Speaker, Session IIA4
WENDY CARTER, Speaker, Session IIA5
WILLIS MADDREY, Speaker, Session IIB1

P-R-O-G-R-A-M

P-R-O-C-E-E-D-I-N-G-S

Opening Remarks 8:03 a.m.

MS. PAULS: Good morning, everybody. We are going to get started. My name is Lana Pauls, and I want to welcome everybody. We are very excited that you are able to join us today. We did not get the 70-degree weather yesterday. Today 65 will be high in the Washington, D.C. area.

We have bathrooms out that direction to the left (pointing) and out that direction to the right (pointing), if you need to take a break. We will be having lunch at exactly 12 o'clock. In addition to that, for those of you who did not get the message, there is an optional evening session this evening that is going to be chaired by Drs. Watkins, Senior, Avigan, and Merz. It will be an open meeting to discuss the formation of a Liver Safety Research Consortium, to be held in this room after the reception at 8 p.m. Again, that is an optional discussion; you did not have to RSVP for it. But, if you would like to join us for a concern in regard to that, you are more than welcome.

First, I would like to welcome Dr. Solomon Iyasu to make some opening remarks. Solomon?

Iyasu photo, biosketch (no abstract, no slides)

DR. IYASU: Good morning. I am honored to have been asked to make some brief remarks. We are delighted to have you here to participate in the 15th Annual Drug-Induced Liver Injury Conference. On behalf of the FDA, I welcome you to this important conference. This annual meeting provides an opportunity to renew contacts and discuss problems of mutual interest in the detection, evaluation, and prevention of drug-induced liver injury. It is a time for us to learn from each other and, also to increase awareness about the problem of drug-induced liver injury.

Over the last 15 to 20 years, there have been significant advances in our understanding of drug-induced liver injury. As you know, the approval of bromfenac and troglitazone in 1997,

drugs that were subsequently withdrawn from the market because of serious and sometimes fatal liver toxicities, really marked the turning point for the FDA in terms of how we evaluate pre-market signals for liver toxicity.

In 1999, the first in a series of meetings was held to consider issues related to drug toxicity that particularly affect the liver. I think it will not be a stretch to say that no drugs have been approved since 1997 that subsequently had to be withdrawn for drug-induced liver injury. This is because of very robust pre-market evaluation and our increased understanding and appreciation for evaluating toxicity in the clinical development programs.

We have come a long way, but our understanding of DILI is not yet complete. We need better data, standardized data collection methodologies, better analytics, and data management. We need better scales for causality assessment. We need to understand the diagnosis of severe liver injury in the context of preexisting liver disease.

Biomarkers to identify susceptible populations would be good, if that could be possible, I think it would immensely improve our understanding as well as target drugs to people who may not be susceptible to these adverse effects.

This year's conference is built around the theme of the importance of getting it right, how to measure and interpret DILI information and to make the right diagnosis. In looking at the presenter list today, you will see that each person brings particular expertise and experience to the science of DILI. Therefore, prepare yourself to be challenged, excited, inspired by the talks, discussions, and dialogs that we will offer over the next two days.

This conference is co-sponsored by the FDA, the Critical Path Institute (CPI), and the Pharmaceutical Manufacturers of America (PhRMA). It has also been endorsed by the National Institutes of Health and the American Association for the Study of Liver Diseases (AASLD).

Many thanks to all of you for coming. I would like to extend my sincere thanks to the organizers of this meeting, and particularly for the leadership of Dr. Senior in making this annual event one of the most anticipated gatherings for many of us. Therefore, without taking too much of your time, I just want to extend my welcome. I wish you every success in the deliberations and, also, to have a very pleasant stay during these two days. Thank you. (Applause.)

Session IA Moderators and Speakers -1:

MS. PAULS: May I welcome Drs. Dufour and Chalasani to the podium, as well as Drs. Karmen, Prati, and Pollock, we will go ahead and open Session I.

DR. DUFOUR: Good morning, everybody. My name is Bob Dufour. I will be one of the speakers later in this session, but I am also going to start the session off. This particular session is related to how we measure our most common biomarker and how we interpret it, namely, alanine aminotransferase.

Our first speaker on the program this morning is a person who was very instrumental in setting up the measurement of ALT. In fact, he invented it. Dr. Arthur Karmen, back when he was a medical student in New York in the early 1950s, was recruited for a project to develop a method for measuring serum glutamic-oxalacetic transaminase (SGOT), now called AST (aspartate amino-transferase). This was mainly driven by the cardiology community. They were looking for a marker for myocardial infarction. Dr. Karmen, working on this project, initially developed a paper chromatographic method for doing it that took several days, which wasn't really practical. But he then developed a method that could be done in a few minutes in the main laboratory of a hospital. His new method was published in January of 1955, although a briefer version was originally published in September 1954, for measuring both SGOT and SGPT (now called AST and ALT).

Interestingly, one of our organizers for this meeting, Dr. John Senior, read the work of Dr.

Karmen in the January 1955 issue of the Journal of Clinical Investigation, and was motivated when he was an intern in Philadelphia to request that he be allowed to run these tests at night to help make the diagnosis of myocardial infarction. So, there is a lot that ties our speakers together today.

Dr. Karmen went on to have an illustrious career as a physician and laboratory director and is now retired, but has come to speak to us today on how ALT and other enzymes could be measured in serum, although pretty much the same methods are still used. So, Dr. Karmen?

Karmen photo, biosketch, abstract and links

AK#1: When I told my wife about coming to this meeting, she said, "Why?" And I said, "Because I'm a historical figure and that's what they think of me." (Laughter.) And she said, "I've been saying that about you for years." (Laughter.) What can I say? I've had some big birthdays in recent times. I have an illustrious family. I have two grandsons, one of whom is becoming a star

for music at the Kennedy Center. He is a second-year college student and he won a contest for arranging and playing music. So, he is going to play in the Kennedy Center here in Washington. And I think that most of our immediate family is about to come for it, and we're having a family get-together. Anyway, I have a brief time to tell you stories. I thought what I shall do is to tell you a little bit about how this all started.

AK#2: I was a second-year medical student at New York University, and the administration of the school decided that it would be better to have medical students training on the wards of Bellevue Hospital work all year round. We thought it was mostly because we did the lab work as students, the clinical lab work. When the students weren't there, the house staff had to do the tests, and they didn't like it. So, to make the house staff opportunities a little more attractive, they changed the schedule so the students had to do that. With the extra time that we would be spending in school, they permitted us to have electives, which meant that we could

study with almost any department we wanted to work with. But this was my first opportunity to work with anything clinical, because, before that, we did all basic sciences. We looked for clinical research opportunities.

I was introduced to Dr. Felix Wroblewski, an internist caring for my uncle, who had been treated by him after surgery for cancer of pancreas, and he had done very well for my uncle. So, I had dinner with Felix Wroblewski and I asked him if he could help me get an opportunity to do some research. He was interested in trying something different at Sloan Kettering Institute where he worked, and he asked me, since he was not able to get me any other opportunity, whether I would like to spend the summer working with him.

AK#3: I did, and one of the ideas he thought about at that time was developing a test, a chemical test, for diagnosing myocardial infarction. Now the question was: where do I start in this? I really had no idea. I knew a little about myocardial infarction from what I learned in the

pathology course in the previous year, but that wasn't much. I looked up myocardial infarction in the famous Harrison textbook of medicine. The only information I could find there said that the laboratory was not involved in diagnosis of myocardial infarction. That was written by the man who was the editor and the major author of the standard textbook of hematology. So, I looked at it and was not able to find any guidance at all. Felix Wroblewski introduced me to the library of Sloan Kettering Institute. One of the textbooks that I found there was a book called Enzymology by Sumner and Northrop. It had a list, almost the kind of thing you might get from today's computers. But I sat with it one afternoon and saw a lot of enzymes listed. We had had some of that introduced to us in biochemistry, but not very much.

AK#4: I spent some time reading that book and finally found on one page that there was an enzyme that was said to be richer in heart muscle than anything else, transaminase. And I said, I wonder if that would be a clue to what might be a good

indicator? Wroblewski told me, "Just look up what the normal range is and set it up." So, I looked up the normal range, but nobody had mentioned about measuring transaminase in blood.

It turned out that somebody else had come later and said, "What is it that you think you're measuring here, because I have read in many places that there is no transaminase activity in blood?" So, I said, "I guess I didn't read the textbook that said there wasn't any there."

I tried to set up measuring transaminase. We had one problem that, when I tried to prepare this talk, was one that I never really emphasized in anything I had written about it. The main problem of what I had to work on was that we had no money. (Laughter.)

So, what I had to do was to have a minimum of expense. Very good. I had no salary and nothing else in the way of income, and my family was not wealthy. But I was paying my tuition well enough with the help of scholarships.

AK#5: What happened? I saw that the transaminase, in particular, had a concentration in heart muscle that was higher than in most other places. When I spoke to one of the biochemistry friends that I had from the medical school, he said "If you want to measure an enzyme in blood, I think you ought to start with something to the liver." But Wroblewski and his senior at Memorial Sloan Kettering, Dr. John LaDue, told me that what the world really does not need is another liver test. (Laughter.)

And that was the way they taught it in those days. When I was invited to this meeting and saw the title and all that, it was amazing that you are all so smart about what was important and what wasn't important. But that is where I came on the scene back then.

I looked for methods that had been used by other people to measure transaminase and transaminase activity. There was one that used a Beckman spectrophotometer. I found out very quickly that nobody in Sloan Kettering would take

a medical student who had been a student someplace else and entrust them with this very expensive instrument.

So I learned about the cost of things. When I looked at the price of a Beckman spectrophotometer, it was something like $2200, and there was no computer to put with it. There wasn't any such thing. For manually operated and interpreted equipment, you had to make do with no computers, no data reduction, or anything like that. The cost of a spectrophotometer was $2200 in those days. The Oldsmobile that I wanted to buy also cost $2200, though I think it went up, in the next two years, to $2600. So, you can get some idea about how long ago this was.

AK#6: Anyway, I found one paper by a graduate student and a professor from Texas, where they had a method for measuring transaminase distribution in different tissues, using paper chromatography. I hadn't heard about that, either. I had been a chemistry major, but my chemistry majoring in undergraduate school was something like four

years, five years before. So, I looked at it and asked a couple of other people for advice. They said that they had a couple of people around who know about it, know how to do it. And I was introduced to the use of paper chromatography, which I presume that most of you either are very familiar with or can find somebody quickly who can tell you about it. It became a very useful tool. I ended up meeting the major inventor of chromatography who had won the Nobel Prize for it, A.J.P. Martin.

Anyway, what it involved was incubating whatever tissue you wanted to test with a small amount of material that you wanted to test -- for transaminases, with an amino acid like glutamic acid and a keto acid, to which it would give its nitrogen group. So, that was the transaminase idea. And it looked like I could borrow some of the reagents for the amino acids and buy some, because they finally gave me a small stipend. I think it was about $50, to start to buy some pipettes and do all that. The rest of it, I got the Bronx way

of obtaining financial support. I stole it. (Laughter.)

AK#7: So, I had pipettes that I could get from different laboratories, particularly if I worked at night. It turned out that the Sloan Kettering Institute, for those of you who are familiar with it, used to have an entrance on 68th Street in Manhattan between First Avenue and York Avenue. It was a very classy, little office that you could use as an alternative entrance. The main entrance for the patients was on York Avenue, and the entrance to James Ewing Hospital, which was City Hospital at that time, was on the other main street. But I used to go in and out at the entrance on 68th Street to Sloan Kettering Institute. It was a very solemn, little place with marble walls but no place where you could sit down. But you could walk through the two hospitals from there. There was a plaque on the wall that said, "Within these walls a few work unceasingly, that many may live." So, that struck me. That is very nice.

AK#8: I found out that when I walked through the walls in evenings, I never saw anybody working at eight or nine o'clock at night. So, I figured if there were a few that worked unceasingly, there were damned few. And that was the end of that. (Laughter.)

AK#9: Anyway, I set up a method by which I chemically separated amino acids. For my first experiment with a blood sample, I didn't have the sense to try to measure it with some liver extract or something like that. So, my first experiment with a blood sample, I separated the material by paper chromatography and looked at the appropriate spot for the second amino acid, which was probably aspartic, as I remember it. And when I looked at it, there was a faint increase in color in an anhydrant spot, more than there had been in the control that I ran right next to it without the incubation. So, it turned out I believed that it was there, but I couldn't figure out what you did to increase the sensitivity of the method.

Finally, it dawned on me that, if it is an enzyme reaction, maybe I could do it better by incubating the possible enzyme with the reagents for a longer time period. I did it overnight. I found out that most people liked the overnight incubations because it meant you set something up at five o'clock, went home, and 16 hours later you came to work at 9:00 in the morning.

And sure enough, there was a nice, big, blue spot detected by the anhydrant, showing that there was transaminase activity in normal blood. And I started working on it. I bought a fish tank with the help of my father, and we did most of the work, made a contraption to hold butter dishes, and we did mass production chromatography. I was able to measure transaminase in blood.

AK#10: For this meeting, all I would add was that I measured the enzyme that was rich in heart muscle. There was a second one. It was uncertain whether it was the same enzyme as was measured with glutamic oxaloacetic transaminase, or one called glutamic pyruvic transaminase.

I bought some alanine and tried it, and, sure enough, it gave almost the same kind of activity in the blood. So, I started to measure both alanine and glutamic transaminases in all my specimens, done at the same time.

AK#11: I got no enthusiasm from either Dr. Wroblewski -- he didn't discourage me from it, but no enthusiasm -- or from the Chief of the Department. Again, he repeated the story that I apparently he had grown up with, that everybody agreed that there was no need for another liver function test.

And it turned out that I would attend this meeting today, and I ought to probably stop my discussion about this with the dictum that they gave me about how the world really didn't need another liver function test, something they all subscribed to. They said they were not going to allow me to use this in the report that I made on my work "Because we know that nobody needs another liver function test, another liver indicator."

I guess that I've told you enough stories. (Laughter.) I have prepared an abstract that tells some of this, and you can read about it.

AK#12: I ended up with a long career, not as illustrious as I might have liked, looking back, but it worked out pretty well. I retired as the Chairman of the Department of Laboratory Medicine at Albert Einstein College of Medicine and the lead director at several hospitals, which is not the way you might want to do it, if you planned for it.

I learned about taxi services. I learned how to speak Spanish mostly and occasionally Russian, from some of our patients. I would not give up the opportunity to do that for anything. I have a grandson that wants to be a physician, and I am rooting him on.

I am getting a signal from my friend Senior. He is pointing at his watch and going like that (pointing to his head -- Laughter) So, either I'm talking too long or I'm somewhat insane. Usually, insanity is this way (indicating -- Laughter)

Thank you. (Applause)

Session IA Moderators and Speakers - 2:

DR. CHALASANI: Thank you. That was wonderful. Our next speaker is Dr. Daniele Prati, who comes from Lecco, Italy. His biography is listed in the syllabus, so I am not going to repeat it. His 2002 landmark paper, in the Annals of Internal Medicine, I think has been cited close to a thousand times. He is going to talk about what is a normal ALT level.

Prati photo, biosketch, abstract

DP#1: Thank you very much. I wish to thank the organizers, particularly John Senior, for inviting me and giving me the honor to speak after this distinguished scientist, after Professor Karmen and before Professor Dufour

DP#2: My talk today will introduce you to what is normal for transaminase. I think that this has some relevance because, as you know from the FDA definition of DILI, the upper normal limit of liver enzymes, not only for ALT, but also for other liver

enzymes, is still important for the definition of disease.

DP#3: And the upper limit of normal still is used in the even more recent definition of the disease, like these definitions coming from Spain in which they try to renew the composite algorithm of the Hy's Law to predict the outcome of acute liver failure in patients with drug-induced liver injury. This paper appeared last year.

DP#4: So, the definition of normal still has relevance. In this talk today I would like to introduce some concepts about normality and some considerations about the definition of normal values. And then, next, I will speak to you more in detail about the cutoffs for liver transaminase.

When we talk about normal individuality, we think that, on one side, normal is taken as diagnosed free from a given disease, but on the other side we define diseases as those that are different from normal individuals. And this is an example of circular reasoning.

DP#5: I would like to draw your attention to the difficulties of defining normality. I would like to start with an idea of Sigmund Freud that, for the first time, more than a century ago, challenged the idea of normality. These are words from his late work, Analysis Terminable and Interminable. He said that every normal person, in fact, is only normal on the average, and this is a statistical concept, and that his ego approximates to that of the psychotic at some point or other, and to a greater or lesser extent. And this means that any definition requires individualization.

DP#6: Also, a seminal paper about the concept of normality came from E. A. Murphy, published in 1972, and shows you some different definitions of normality from different disciplines. You can see here that in clinical medicine we identify normality as healthiness, but we need to define it from a probabilistic approach. And so, we need statistics.

DP#7: So, we use the Gaussian distribution or other distributions. The Gaussian distributions, I remind you, mean that we define as normal all the values that fall within the Gaussian curve.

DR#8: And to deal with more complex distributions like this one that is, as you can see, skewed on the right. So, it means that there is some contribution from subjects with subclinical disease. It means that we use the 95-percent percentile of the normal population.

DP#9: So, in any of these definitions, anyway, we have some drawbacks. The most important drawbacks were underlined by Sackett and Haynes several years ago saying that the probability of an individual to be called normal depends on the number of tests we require. So, if we order one test, the probability of testing to be normal is 95 percent, but if we order 20 tests, the probability of testing normal is only 35 percent. It is only 0.6 percent when you order 100 tests. So, this is due to the statistical definition of normality.

DP#10: The concept of reference values was, then introduced in place of the ambiguous term of "normal values," and the term of "reference values" would apply to any type of reference individuals, whether healthy or unhealthy, assuming that the reference values were properly qualified.

DP#11: And so, the idea was to change terminology and not define anymore the normal range, but talk, rather, about reference ranges. When these ranges have been calculated in an apparently healthy population, we should use the healthy-related reference ranges, or healthy ranges.

DP#12: And this also because having just a normal value can be quite a simplistic approach. As you can see here, we can see that the normal limit is the same throughout all the natural history of the disease, while probably the next table of a reference population in different phases of the disease and in different steps, like prevention, screening, and diagnosis, and that affect the survey. And so, we could use different

populations like healthy individuals, those with recent complications, and those who have responded or been cured from the disease.

DP#13: In this regard, I would like to draw your attention to this paper published by The Cochrane Hepato-Biliary Group last year. The senior author was a colleague from my group. As you can see here, probably the slide is not very clear, but we can divide the diagnostic process into phases. The architecture of diagnosis goes in phases, phases that are very similar to the clinical trial phases.

DP#14: And so, starting from Phase Zero to Phase 4, we have different phases in which we define threshold among individuals at the beginning, but later on, also, defining with ratios on the basis of the outcomes. This process, any diagnostic process, deals with certain cutoffs or uses cutoffs for comparing them to predict different outcomes.

DP#15: So, the possible approaches are the healthy ranges, the diagnostic thresholds in subjects with the disease. The thresholds with

complications, to confirm complications, outcome studies. These are morbidity and mortality among both the healthy subjects and among the subjects with some complications, but, also, the idea of finding individual reference ranges, which is a very complicated task, meaning that the reference subject is the subject himself when he was healthy or uncomplicated. These are, of course, very complicated approaches.

DP#16: So, in any case, if we deal with a continuous variable and we want to predict the risk of the disease, it is better to speak about the different levels of specificity and sensitivity that drive us to define the different thresholds, in other words, challenging the idea itself of normality.

DP#17: For any distribution, moving the threshold from one level to another means that we decrease the number of false-negatives but we increase the number of false-positives. This can be done when we are approaching a disease. For example, it has particular characteristics, as

being easily treatable with a drug, with a new dose of drug, for example, that has very low side effects. So, in this case we can probably tolerate a higher number of false-positive results while, before that, we could not tolerate it.

DP#18: So, for liver injury and the normal range, we can apply different approaches, as these approaches have been all used to define the cutoffs for alanine aminotransferase. The low-risk population is what I showed you in the distribution among the normal individuals. The mortality definition, predicting mortality for chronic liver disease, differs from the values observed in individuals cured from liver disease.

DP#19: The first step, the one from low-risk populations, means that we have to exclude the patients with a high risk of liver disease, including those with hepatitis B, hepatitis C, alcoholic liver disease, and fatty liver disease. So, even in this case, having different ratios for different clinical outcomes, it would probably be more appropriate.

DP#20: When we started work in the early 2000s, we had a laboratory marker, serum ALT, that was widely used, but had very poor value for diagnosing liver disease. Technical standardization was lacking. It was improperly used as a marker of fibrosis, and advanced liver disease. And the reference values were not updated for a long time.

DP#21: So, we found that reference populations proposed by laboratories as “normal” contain a substantial proportion of people with subclinical, undiagnosed liver disease. Our idea was to take a symptom-free population, a blood donor in this case, to exclude viral hepatitis carriers and to exclude also those at risk for steatosis, those with high lipids, high glucose levels, overweight individuals, and heavy drinkers, and calculate their reference interval.

DP#22: In this way, we were able to show that we could lower the upper limit of “normal” from 40 units for males and 30 units for females that were average at that time, to 30 units for males and 19

units for females. This work was published, as I said, in the Annals of Internal Medicine.

DP#23: Another approach that followed this was proposed by a Korean group that used receiver operating curves for ALT to identify people at risk from death of liver disease, a mortality approach. And this approach, surprisingly, gave similar results. So, as you can see here, 30 units per liter was identified on the basis of survival data.

DP#24: We also tried to apply the definition, calculating the distributions in people that had been cured from liver disease from hepatitis C. Again, in different labs we had a 25-percent reduction, meaning that from 40 we went to 30. So, this summarizes these three approaches published between 2002 and 2006.

DP#25: More recently, another group calculated from living donors, living liver donors, the distribution of ALT, taking into account not only the metabolic parameters, but also the histological data from these donors. They found that among the 600 individuals selected in

the material published in our Annals of Internal Medicine paper, the healthy values were 33 units for men and 25 units for women.

DP#26: And this is a slide from a recent review that shows you that almost every paper published so far shows higher levels in males, lower in females, higher levels in those whose metabolic features are not considered and in a group of those who have been cleaned up by the presence of patients at risk for liver disease. Also, this work has been done in children, almost five years ago. This is an American study showing that the 95 percentile in children with healthy weight, was 25 for boys and 22 for girls. If you wanted to review in depth all this data, you can read it. It is hard because it has been published by another Italian group and shows in very great detail almost all the studies that have been done during the last decade.

DP#27: So, is everything definitely set now? I would say no, because the upper reference limits in different labs are still variable, and expressing the results that are done as upper

normal limits does not solve the problem. We do not know what might be the impact of the introducing standardized information on chemistry based on the reference ranges. And there has been no clinical validation of these ranges so far. And I would also say that most clinicians don't even know that this has been changed over time. So, I think that this would be the topic of Professor Dufour after this presentation.

In addition, we almost thought up until now about ALT, but the upper limit cutoffs for SD phosphatase and bilirubin, among all the others, have not been so far very well-identified, not with the same analytical attention. The cutoff varies for drug-induced liver injury. They are not based on prospective studies. This is a major problem because this implies that it would be difficult to define upper thresholds based on outcomes. And so, of course, we need new studies.

DP#28: Just to summarize the message that I want to leave you with today, several studies conducted during the last decade suggest that the

healthy ranges for ALT can provisionally be set around 30 units for males and 20 or 25 for females.

But we have also to think that a simplistic approach of using a universal cutoff for all clinical situations should be abandoned. Results should be interpreted more flexibly, taking into account the scope of the tests and especially the patients' characteristics in terms of gender, age, history, and also other risk factors.

In addition, the definition of the ALT ranges, as we have seen, is only the first step in evidence-based diagnostic research, and a cutoff should be ideally identified on the basis of prospective studies based on clinical outcomes.

Setting the appropriate cutoffs needs close cooperation between the pathologists, clinical pathologists, statisticians, and the regulatory institutions. This is very important because it has not always been the case. So, sometimes hepatologists create their own thresholds and clinical pathologists may establish another. It

is important to have multidisciplinary work. Much work is still needed.

To finish with an idea that comes from a very important but non-medical book published almost a century ago, let me quote from a book that is called The Confessions of Zeno, by Italo Svevo. The book focused on the relationship between health and disease. Using as an example thyroid disease, that fits well also for liver disease, it says "All organisms extend along a line. At one end is Basedow's disease, which implies the generous, mad consumption of vital force at a precipitous pace, the pounding of an uncurbed heart. At the other end are the organisms depressed through organic avarice, destined to die of a disease that would appear to be exhaustion, but which is, on the contrary, sloth. The golden mean between the two diseases is found in the center and is improperly defined as health, which is only a way station. In the middle are those who have either incipient goiter or incipient myxedema, and along the entire line, in all mankind, absolute health is missing."

So, thank you very much. (Applause.)

Session IA Moderators and Speakers - 3:

DR. CHALASANI: Thank you. Very nice. Our next speaker is Professor Robert Dufour, who will continue discussion on ALT. He is currently a consultant for pathology and hepatology at the Washington VA Medical Center. He is an emeritus professor of pathology at George Washington.

Dufour photo, biosketch and abstract

RD#1: 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.

RD#2: So, if we could have the next slide, please? Unfortunately, I have no relevant conflicts to disclose. (Laughter.)

RD#3: 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.

RD#4: 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.

RD#5: 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. 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.

RD#6: Now in the United States the major provider for proficiency testing materials it the College of American Pathologists. 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.

RD#7: `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.

RD#8: 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".

RD#9: 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.

RD#10: 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.

RD#11: 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.

RD#12: 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.

RD#13: 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.

RD#14: 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.

RD#15: 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.

RD#16: 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. They developed partnerships with the CDC, with the NIH, with the National Institute for Standards and Technology, and a number of laboratory groups, to try to work with manufacturers and improve the repeatability of results of hormones among different laboratories.

RD#17: 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.

RD#18: 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.

RD#19: 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. (Applause.)

Session IA Moderators and Speakers - 4:

DR. CHALASANI: Great talk. Our next speaker is Dr. Nira Pollock, who is going to talk about a very interesting point-of-care testing for ALT. She is at Boston Children's Hospital.

Pollock photo, biosketch and abstract

NP#1: Thank you. As a disclaimer, I am not a DILI expert. I am a diagnostic developer. But I have been working on this project for a number of years in collaboration with Diagnostics for All and the Whitesides Laboratory in Boston. And I think it fits very nicely into this session.

I am going to move fairly quickly through some of the older published work to get to some of the newer unpublished work, which I think is quite relevant to what we have been discussing.

NP#2: The motivation for this, for the development of a paper-based finger-stick transaminase test is pretty clear to this group, right? So, blood tests for monitoring liver status are standard in developed nations, but in developing nations are often unavailable. We know that standard-of-care testing requires venipuncture, centrifugation, large automated platform, most typically, et cetera. In particular, the centralization of testing in centralized laboratories can introduce major delays in results return.

NP#3: This group knows very well that testing for transaminase is particularly important for patients in resource-limited settings, so particularly patients on HIV and TB medications. I am sure you are quite familiar with the

percentages of DILI in these patients and the relevance of testing in these settings.

NP#4; So, the motivation for us was that a cheap and accurate and point-of-care test for measurement of, quote, "liver function" would probably have a very dramatic effect on patient care in the developing world, in particular.

NP#5: I started working with Dr. Whitesides and the Harvard Chemistry Department and Diagnostics for All, which is a nonprofit in Cambridge, to develop this test a number of years ago. And importantly, the test is based on a very interesting technology called patterned paper, which originally came out of the Whitesides Laboratory and, then, was taken further by DFA. So, we started working together in 2009 to further develop this test and started getting some grant support to do that.

NP#6: This is a schematic of the test itself. What you can see in the upper lefthand panel, it shows the design, the schema of these types of tests. So, we have layers of patterned paper.

The way you pattern paper is by taking a wax-based printer and a heat source and patterning hydrophobic channels, which then guide the wicking of fluid through the paper.

When you have vertically-stacked layers of patterned paper, you can start with a drop of blood, add it to the top. A plasma separation membrane filters out blood cells and, then, plasma continues, wicks into the paper, separates into different zones, picks up dried reagents necessary for zone-specific chemistry, continues down to the next layer, picks up new reagents. And then, you can read the test by evaluating each zone separately. So, you have different chemistry going on in different zones.

In the center you can see this is the early version of the test where we had AST and ALT on the same test. This test is the size of a postage stamp and is meant to be very low-cost, than 10 cents per test.

You can see that you have color change in the AST and ALT test zones that you can read like a pH

strip. So, with 15 minutes of development, you see an increase in color correlating to the increase in the transaminase concentration.

We also have control zones which, then, are interpreted as to whether the test is valid or invalid. And I will come back to that in a later iteration of the test.

NP#7: So, we started with analytical performance evaluation, as you would for any test. We established that the test was linear across the clinical development range. We looked at limits of detection for this early prototype. We looked at repeatability, everything you would do in the beginning proof-of-principle stage of a test like this.

NP#8: We went, then, to a preclinical evaluation which we did in my hospital, Beth Israel Deaconess, where we looked at clinical venipuncture samples that had been drawn from patients, paired samples where the patient had whole blood and serum drawn at the same time. The serum had been tested on the automated platform.

We tested serum and whole blood on the paper-based platform. We read the test 15 minutes later with three people reading it separately to see if they got the same answer, et cetera. And so, in all of these studies, we are comparing results of the paper test to the results of an automated reference benchmark.

NP#9: Importantly, we decided to sort of have bins of results, which were matched to TB, in particular, TV and HIV treatment guidelines, which, as we know, use cutoffs like less than 3X upper limit of normal, 3 to 5X and greater than 5X. And we can discuss -- maybe that is not sort of the right way to define normal, but that is how we decided to set up this test.

NP#9: So, this is a schema from the paper, which you can look at if you are interested, which just basically showed that in serum and in whole blood for both ALT and AST we had greater-than-90-percent bin placement accuracy for this test, meaning that the readers put the test put the test in the same bin, less than 3X, 3 to

5X, greater than 5X, as the automated method at least 90 percent of the time.

NP#10: We did additional evaluation to understand assay interference from relevant physiologic markers, bilirubin, cholesterol, et cetera. The takehome from this for us was that this test performed pretty well in a diverse patient population. Many of these specimens were from critically-ill patients. And we were, then, encouraged to move forward into a field study in a target patient population.

NP#11: We then moved into a study in Vietnam, which we did in collaboration with PATH and the Hospital for Tropical Diseases in Vietnam and the Harvard AIDS Initiative there.

NP#12: The objectives of this study were to really evaluate whether this test could be used. Was it feasible in a target population with target operators? So, we were interested in feasibility, inter-operator and lot-to-lot variability, device failure rate, and then, secondarily, device accuracy. This was really to understand whether

people could perform and read this test and get the same answer.

NP#13: So, in this setting we used a new iteration of the test, which is a three-zone test. For reasons which I won't go into now, we decided to focus exclusively on ALT. You can see that the device has become a three-zone test with an ALT zone at the top and two control zones, a negative and a positive control. THis test was a finger-stick study where we used a capillary tube to collect blood from the finger stick and apply it to the device. It was operated by two nurses in the clinic who, then, each read the test independently, and we compared their results to each other.

NP#14: This is an example of valid and invalid test results. So, the two panels on the left are valid test results and the four on the right show different ways to get an invalid. You can see in the negative control zone that it needs to turn yellow; it needs to turn from white to yellow if enough blood is there. If it doesn't, there wasn't enough blood. You can see in the negative control

zone in panel 4 that, if the negative control turns red, there hemolysis and that is an invalid result. We don't want hemolysis because, then, we can't read the test the way we want to. And there is a positive control zone where, if it does not turn red, we know that the test is not activated properly. So, any invalid result, if any of these types, invalidates the entire test.

NP#15: The study design was to go training first, then a pilot phase, and then a large-scale evaluation of 600 subjects. These were subjects from a very operative HIV clinic in this hospital. We performed the paper-based test on finger-stick blood, as I said, and read it with two people. We compared the results to gold standard, which in that hospital was a Roche Cobas platform using plasma.

NP#16: So, the results: we enrolled 600 subjects, and we used two device lots, which ends up being relevant throughout the study.

NP#17: Now the main take-home message from this study for us was that the interoperator

variability was low, meaning that the two operators got the same answer almost all the time. So, in terms of determining valid versus invalid, the two operators, Nurse 1 and Nurse 2, agreed on valid versus invalid greater than 90 percent of the time. And they agreed on bin placement almost 96 percent of the time. So, the two operators reading this test visually got the same answer almost all the time. That was really the main take-home for us, was that people could read this test.

NP#18: This is a pictorial example of the two nurses' results correlated against each other, and you can see that they match very well.

NP#19: We looked at device failure rates and learned something very important for device optimization, which was that our two lots performed differently. We saw 21 percent hemolysis rates in one lot and less than 2 percent in the other. We learned that the plasma membrane itself had a problem and needed to be resourced, that we were getting a lot of hemolysis from one particular

batch of this plasma-separation membrane. Otherwise, we saw very low rates of invalids.

NP#20: This is an example of the ALT results themselves. You can see that overall bin placement accuracy here was around 85 percent for both Nurse 1 and Nurse 2, which led us to understand that we needed to do more work to optimize the accuracy of the results, but we understood that people could actually perform and read this test.

NP#21: Operational assessment, we learned that people could do it. We didn't know exactly how much training people needed because we did it a lot, but we could see that they were able to train each other. They were comfortable using it. We learned that people could apply blood to the test properly, that we didn't have invalids due to insufficient sample volume, that we could do this with finger sticks. We learned everything like that about feasibility.

NP#22: So, the conclusions were that this was a successful field study in a target environment,

tests performed by the people we wanted to perform it. Inter-reader agreement was excellent.

But we had areas for improvement, lot-to-lot variability. Bin placement accuracy had to be better. We needed to get rid of this particular lot of plasma membrane.

NP#23: So, the questions at this stage, which I think are quite interesting to me as a diagnostic developer, this test is the first device like this. There are a lot of people doing paper-based microfluidics, but this is really the first test of this type of technology to come this far towards clinical use. So, there really aren't clear precedents for performance standards. How accurate should a device like this be in order to be clinically-useful? Does it have to exactly match an automated platform? Does point-of-care utility override that? Does cost override that? There are all sorts of really interesting questions about how a device like this could or should be used.

NP#25: We then went to device reoptimization. We got a new plasma-separation membrane. The assay chemistry was reformulated by DFA to improve the readout, particularly in that 120-units-per-liter range. And it was recalibrated against a new automated reference standard, which is the Abaxis Piccolo, which you may be familiar with.

So now, this is the unpublished data. This is the clinical validation of the optimized paper-based test, so the ALT test, which we did over the past year, also, in Beth Israel Deaconess and the Liver Center and ID clinics. We enrolled 96 patients that had a range of ALT, all ambulatory. They had a varying range of underlying liver disease. We did finger-stick testing. We had 1-percent performing the test and, then, we had a reader from DFA, from Diagnostics for All, reading the tests. They were blinded to the baseline results for that individual. They never met the patient. And then, we captured the image of the resulted test after it was read with a cell phone

camera. Actually, we used two. We used the fancy camera and we used the cheap camera to mimic what might be available in the developing world. And we texted those results to an offsite reader, who would then read them on a computer screen to see if you could do that, if we could take a picture and have someone read the picture and get the same answer as reading it in real time. The patients were all going for venipuncture anyway for clinical testing of serum. We, then, captured the discarded serum when it was ready to be thrown out, brought it to the lab, tested it on the paper test and tested it on the Piccolo. Okay? So, a lot of comparisons and we learned a lot.

NP#26: So, what you see here, first, are the correlations between the results of these different specimen types and these different platforms. So, in the upper left, we see the finger-stick DFA result versus the serum DFA result. What you see is very good correlation, but an interesting finding, which is that the finger-stick results were systematically below the

serum results when tested on exactly the same platform. Okay? If you look to the upper right, you see that the finger-stick DFA results, again, versus the Abaxis results for the serum were again systematically below, the finger stick was systematically below serum results. So, for the same individual, a mismatch between the finger-stick result and the serum result, but yet a systematic one.

If you look in the lower left, you see that you, when you test the serum from one patient on the DFA device versus the serum on the Abaxis, you get very tight correlation. So, the DFA paper test was matching the automated Piccolo test very, very well.

In the lower right you can see that the serum tested on Abaxis was systematically below the result of the serum tested on the Roche, which was our clinical method. So, a systematic difference between the results of two automated platforms testing the same serum sample.

NP#27: This is just another way to interpret these types of comparisons. This is Bland-Altman analysis, and this is set up in the same way. You can see that, when you compare finger-stick values to serum values for the top two panels, you see a negative bias. So, the finger-stick value is coming out below the serum value, whether you are testing that on paper or on Piccolo.

NP#28: And then, on the lower two panels, the same comparison, you see that serum tested on paper and serum tested on the Piccolo match very, very well. And then, you see in the lower right that serum tested on the Piccolo is systematically below serum tested on the Roche. So, we learned a lot from this.

To conclude our data, we saw that when we had people read the device, the texted images of these devices on a computer screen, they got the same answers as those reading it at point-of-care. So, we learned that, whether we used the cheap camera or the fancy camera, you could do this and you could get the same answer from a trained reader reading

it offsite. So, the conclusions from this study for us were that the paper ALT test was highly accurate for serum testing. In fact, it matched the reference method that we used to optimize the test, the Abaxis. Better than the two reference methods, Roche and Abaxis matched each other, which was interesting to us. We saw that Abaxis was systematically 9 percent below Roche, and we know that this has been seen in proficiency testing.

NP#29: Very interestingly and unexpected to us, we saw a systematic difference between ALT values measured in finger stick versus paired serum, and that finger-stick values were systematically lower than serum values for the same person. So, we went back, of course, to the literature, trying to understand whether there was precedent for this, looked at the literature for the few finger-stick ALT tests that are out there. So, there are some automated finger-stick platforms which you may have used yourself. There is the Roche Reflotron. There is the Alere Cholestech. What we were surprised to learn, when

we looked into their literature, was that when those tests were validated, they were actually never validated with finger-stick samples from patients with truly elevated ALT. The highest ALT that we saw in any of the data, in the package inserts and data out there online, and so on, was 65. So, they never tested people with ALT values higher than that. So, it is possible that, if there is this true systematic difference, which we found there was, between finger stick and serum, they would never have seen that because they never tested people with truly elevated ALT, which is pretty interesting to us. You could use confirmation by others in this room. We ended up calculating a correction factor for our device, which would, then, correct the finger-stick data to have it match the serum data, because we know from our data that the serum tested on paper matches the automated platform extremely well. So, by correcting the read guide, we can, then, have a finger-stick value that matches an automated test.

Okay? We also learned that remote reading is feasible.

NP#30: So, to conclude, the next steps: FDA discussions are currently in progress. There is interest in demonstrating safety and effectiveness both in the point-of-care and, also, potentially the home-use setting. The laboratory is gearing up for production under GMP, and we have done our writing-up a cost-effectiveness analysis to understand the utility of this device in the settings for which it has been targeted.

NP31: So, here are my collaborators. And I will stop there. Thank you. (Applause.)

Session IA1 Discussion:

DR. CHALASANI: Thank you. That was a wonderful talk. I think we have about 20 minutes for discussions. If you have any questions, please come to the microphone. I will start with a question to Dr. Prati. Do you think that your suggestion that we have different reference ranges for different clinical scenarios, is practical? Do you think we could pull it off?

DR. PRATI: Well, I think that this would be probably the best way to comply with the idea of evidence-based medicine. So, we need to prevent outcomes, and we have to base thresholds on outcomes. This is one thing.

The other thing is that probably we still need to know what is the healthy range among a healthy population. This is what I tried to show you in my presentation. This is something that has to be done because this is the first phase of any further reasoning about the use of a test.

I think that in the case of drug-induced liver injury we have to think that the upper limits that

we use are for some criteria, for example, the FDA criteria that are based on the Hy's Law, and Hy's Law was based on an old definition of what these healthier ranges were.

So, I think that we have to consider all this, put all these things together, and probably for the future try to build cutoffs rather than simple healthy ranges.

DR. CHALASANI: I don't like either what goes on, but the difficulty is that when we are building a reference range, do we know it is linked to poor outcomes, or miss cases of DILI? Because industry, at least the majority, uses central labs. Maybe in clinical practice, having a varying range may be problematic. Even in clinically significant DILI, there are some who are jaundiced, et cetera. I am just wondering, because what you are proposing is quite daunting in terms of achieving practically.

DR. PRATI: So, I am saying that I think that we probably need, first, to have very stable reference ranges. This can only be achieved if we

choose the same rules with regards to the definition of the reference population and when we have very stable tests, as Professor Dufour underlined.

So, when we have this ideal scenario, we need another ideal thing. That is a very large database of DILI cases measured with stable tests. From this database, we can probably find the best cutoff to identify the DILI cases. Because we have to think that we probably need the different cutoffs if you are in the phase of licensing the drugs or if we are just following patients. So, the approach would be probably very different in terms of sensitivity and specificity that we would require to data examination. I hope that it was clear.

DR. CHALASANI: Mark?

DR. AVIGAN: Mark Avigan at the FDA. I have a question for Dr. Pollock about home use and point-of-care use for these devices. You have a colorimetric test which requires color recognition by an observer. So, it turns out that, among

males, maybe more than females, there is a fair number of people who are actually color-blind. There is a question of reliability of the observer in settings where the observers themselves have not been standardized. Have you thought about that with regard to that question?

DR. POLLOCK: Yes, we have definitely thought about it. I am not personally involved in the setup of the FDA clinical studies. We can happily discuss that later. There are those here who are involved. This was part of the rationale for developing alternatives, right? So, you can easily imagine that you could, for a person who was colorblind, either do what we did in the study, which was to take a picture of it and send it somewhere else and have someone else read it, or to have an app. For many of these tests, people are developing apps where you could just take a simple picture and it can interpret the color for you. So, there are a lot of workarounds. But, yes, that is one of the potential failure modes for sure of a test that would need to be read at home.

DR. AVIGAN: Just on that same point -- it is usually important when you are talking about low inter-observer variability. You have an "N" of 2 in your clinic. It’s important, if you are expanding to a very large use with many observers, across people who are elderly and perhaps have other visual problems, or might have CNS issues, that the variability might be much larger than you reckoned for with just two test observers.

DR. POLLOCK: Absolutely. I completely agree. We also don't know. In that study we trained people quite well, right, because we were trying to evaluate feasibility and didn't want that to be a barrier. So, we don't actually know the minimal training requirements and whether someone can simply read the package insert and, then, perform the test properly. So, those are all very important points for the upcoming studies.

DR. CHALASANI: Arie?

DR. REGEV: Arie Regev, Lilly. I have a comment and then a question. First of all, I enjoyed the talks, very nice, very enlightening. But I am

trying to figure out what the approach is to get to a common reference range. What are they trying to solve in the DILI world? Clearly, having a common reference range would be better to compare results in general between different databases. I think that makes a lot of sense. But I heard comments mentioned like, when we have that, we will have a better ability to diagnose DILI. I am trying not to be a party pooper, but I don't see the current world of DILI, how having a 253 ALT versus 272 may solve our problems with either diagnosing DILI, assessing the potential outcome of that ALT elevation, predicting which patient will or will not have drug-induced liver injury, and so on.

DR. CHALASANI: All good points, though I am thinking that if you use a central lab, I think the 25 is the upper limit of normal, meaning if you use the central lab, that may not be an issue. But if you do a trial where upper limit of normal in Richmond is 80 for ALT, whereas it is 45 in Terre Haute -- I doubt it -- but I think that may introduce

some bias. That bias, I guess, would be symmetric between two groups, you know, if it is in a placebo versus an active group. But, nonetheless, I think you are picking up one issue. But, generally speaking, the difficulty in diagnosing DILI is not necessarily because there is an ALT elevation here or there. It is more to do with excluding competing etiologies, whether they were taking other medications or have undiagnosed hep E, hep A , C, or B. Those are more important, I think, in diagnosing currently DILI. So, did I answer your question, Arie?

DR. REGEV: Yes, and I agree. It would be nice to have and better to have a common threshold. I agree completely. But from the point of view of addressing the big issues of DILI that we are facing, I think there is an additional conceptual issue here. I think John Senior spent the last 10 years trying to convince drug companies that ALT levels are really not the issue; that if it is two times the upper limit of normal or three times the upper limit of normal, as long as the bilirubin is

not increased, those are not the real predictors of DILI. And so the discussion of whether the ALT is 43 or 38 I think misses the point of how are we addressing those issues --

DR. CHALASANI: Yes. Once again, Arie, though, I think you are seeing through the lens of a central lab. You truly are. Because I think there is a wide range. You could see that in Indiana in the paper we published, when we looked at all labs, 90-some labs, it really ranged, the upper limit of ALT ranged from 57 to 85. That causes a problem. For example, you have eight times the upper limit of normal, five times. You know, two or three is okay. There I think there is a potential for a problem. I just wonder, though, next to duration, my own understanding is machine-to-machine variability of a given sample is not as much as how these instruments or devices have the reference ranges. So, the next iteration I just wonder, as we set up three times, five times, we just go with the discrete values, like 150, 250, or 350. That is a thought.

DR. PRATI: May I just offer ---- something for our colleague. I think that you really catch the point. I don't think that for the clinical world we need thresholds that are written in stone. But, if we make some limits just for ALT that are needed for deciding whether to go on with the trial or not to go on with a trial, to go on with a patient or not, I think that the least thing, that we need at least the threshold, that the cutoff is decided with the same rules. Otherwise, we will increase the number of false-positive or false-negative results.

DR. DUFOUR: And I think the problem with the approach that you suggested, Dr. Chalasani, is saying 150 or 300, or whatever, that works well if the results are comparable between laboratories, but there is still a fair amount of method-to-method variability. So, there still does need to be improvement of that before we could sample what those cutoffs would be.

DR. CHALASANI: Hans?

DR. TILLMANN: A question for the speaker on point-of-care. Could you do it with an app that automatically gives you more exact numbers? Because an app could differentiate the different shades of red better than an eye could, so that you actually could perhaps come into a 10-percent range of the actual range.

DR. POLLOCK: Yes, so you could, we know from work that is early development work and ongoing also, has involved scanning the device and, then, turning that, the color density, into a number. You're right that, when we have read it visually, we have asked readers to just round to the nearest 10 units per liter. The sort of thought process with the device has always been that we are really sort of -- you can see from the early work with bin placement accuracy that it has really been kind of a --- where in the range are you? What should the management decision be that follows from that range? But, yes, technically, you could convert that. We know that it is a linear relationship.

We know that you could convert the color density into a number. So, yes.

DR.TILLMANN: A colorblind patient who needs to read it would need to send it to someone else, which is more expensive than if you develop an app which is used by a million people.

DR. POLLOCK: Right. Yes. And there are other factors with using apps and what phone and validating the test and the app together versus separating them. And so, there are a number of operational issues. But, yes, to your first question, if you really wanted to get an absolute number from it rather than a rounded number, you could use some kind of an app.

DR. HAIT: Yes, thanks. Will Hait from Janssen. Just one comment and one question. The work that was cited about upper limit of normal AST/ALT levels in children, was Jeff Schwimmer's work from San Diego. It was in children about age 4 to 18. I would like to point out that there have been some studies that looked at younger children and have shown that the upper limit of normal for

younger children in both a population of children who went to women with hepatitis C and, then, followed for about two years in Europe to make sure they didn't have hepatitis C, and most of them did not, those children had AST/ALT levels. It showed that the upper limit of normal for ALT was about twice the upper limit of normal that you and others have shown in adults for both girls and boys. Interestingly enough, the difference between girls and boys didn't become manifest until after one year of age.

We did some studies on a population of 400 kids between the ages of 1 month and 12 months of age who were being studied for gastroesophageal reflux in a number of registry studies. They had baseline tests, even in otherwise normal children, they had baseline AST/ALT. We diagnosed at an upper limit of normal, 95th percentile of about 63 or 64 ALT, with no difference in the first year of age between girls and boys. So, that is my comment. Kids are a little different from the adult values.

The question is for Dr. Pollock. As a pediatrician, I always worry about hemolysis with finger sticks, especially in small infants. We always intuitively think that that may actually raise the finger-stick values of AST and ALT, since there is AST and ALT in red blood cells, higher than serum values that might have been obtained from venipuncture where the blood is flowing more easily. Can you tell me why you think the finger-stick values that you showed were less than the serum values?

DR. POLLOCK: Well, I can't answer the question of why, of why the finger-stick values were systematically low. We have interpreted this as a physiologic difference in concentration in these sample types, in plasma obtained from a finger-stick capillary blood sample versus serum.

As you may know, when you look in the literature for other analytes, you can see plenty of reports of differences between finger stick and serum or finger-stick -- let's say finger-stick and serum values. But they can be in different

directions for different analytes, one higher than the other, one lower than the other.

And so, we didn't really get any great answer from our review of the literature as to why we see this difference. I think it would be excellent if others would be interested in confirming that. Certainly, that is something that will be ongoing.

So, I don't know about why. I know that in terms of hemolysis we are following that very closely. Obviously, in our study we didn't see -- we had no invalids due to hemolysis. And so, that was not a factor in our study, but that is visually-interpreted hemolysis. And anyway, ours was lower. Our finger-stick values were lower than serum. But I don't have an explanation. I would be very interested if anyone in the audience had an idea.

DR. DUFOUR: We have time for a last question.

DR. DUNN: Laura Dunn from the FDA. In full disclosure, I have been involved with the review of this finger-stick device.

But one comment -- and I think Naga said this, but maybe it didn't come out clearly enough -- that an ALT level should not be a reason to withdraw a patient from the drug or stop a trial. The ALT level should just be a trigger to evaluate the patient, and then, the critical decision has to be made to determine whether that patient may or may not have DILI. So, don't think that we think that the ALT level by itself is an appropriate determination of stopping the drug or discontinuing the trial.

My second point I wanted to make was that I have read some literature that showed that blood samples sent from the same patient to several different labs came up with the same value, no matter what the upper limit of normal value was set by the lab itself. So, you questioned that. You said there was a lot of variability. So, I wonder, how much data do we really have on that?

DR. DUFOUR: So, really the data are fairly limited, as I mentioned, on patient samples. Most of the data that exists is using proficiency

samples, which are not necessarily the same as patient samples.

I can tell you, as a clinician in the VA system, we are able to look at results from all other VAs and most DoD facilities. I can tell you for most of my patients that I have seen, no matter what laboratory they were tested in, their ALT results have pretty much been in the same range. So, that has been my clinical experience as well, that it doesn't seem to make a lot of difference.

But there would be a way to evaluate that, by using frozen samples. Now using frozen samples is a problem for ALT because ALT is going to decline in frozen samples. So, it is much more difficult to test patient samples than stabilized samples, and that may be part of why we are seeing that variability.

DR. DUNN: Well, if you use frozen samples, and even if the ALT was decreased, if you could show that different labs that tested it all came out with the same number, you could say that there was uniformity among the results and you could set a

common standard of upper limit of normal in the different populations.

DR. DUFOUR: I will say that, for example, with cholesterol, I didn't go into all the details because that was more for laboratory interest, but there was a big issue with cholesterol, that when we used proficiency samples, that that resulted in a lot of variability that wasn't existent when they used fresh samples. And so, they actually now use fresh patient samples that are sent overnight to prevent them from deteriorating. That has led to a lot better results and a lot better agreement. A similar approach may need to be used for ALT, but, again, there has not been any push for doing this because there has been on interest in trying to reduce the variability that exists right now, if there is some.

DR. CHALASANI: Laura, I think my own take is machine-to-machine variability today, as machines really there are four or five major manufacturers, whether it is the proficiency samples. A couple of small studies were done. I don't think they

were ever published. One was from Colorado. We tried one. When we sent it to different labs, I don't think we see a big difference. The big difference is really coming from the reference ranges, how the manufacturer is setting up.

DR. PRATI: If I may add something, I am not an expert in laboratory medicine. I am a clinician. But, as far as I have read, we have now methods that are the International Federation of Clinical Chemistry methods that are quite stable. And they are able to give you good repeatability, even between different laboratories.

The point is, is everybody using these methods or not? And also, are the reference values updated according to these new methods? And which are the reference populations that are chosen in each laboratory to give you the reference ranges? So, there is still some major source variability, including the other variables such as, for example, the addition of pyridoxal-phosphate to the AST, which is able to give you different readings. If you maintain in any way the idea of correcting the

final results according to the upper limit of normal for each lab, we will still face some variability. Probably this variability we would want to avoid in the future.

DR. CHALASANI: It has been a great discussion, but I think we should break for a coffee and be back here at 10:00. Thanks to the speakers.

(Applause.)

Session IB:Moderators, Speakers -1 10:00 a.m.

DR. DUFOUR: I will ask the speakers for the second session of the morning to come up to the platform and we will get started. In addition to my other co-moderator, Dr. Chalasani, I would like to ask Drs. Seeff, Guo, and Kirby to come up.

Our first speaker of the second session is somebody I have known for many years. I won't say how many, but Leonard and I have worked together at the VA hospital for a number of years. In fact, he is the one that got me working in the Liver Clinic in the first place after we started working on the AASLD's guidelines on liver testing.

Dr. Seeff is now working on drug-induced liver injury. He was formerly with the VA, then with the NIDDK, and then, with the FDA.

He has worked in a number of areas related to liver disease. He is going to talk to us about what the best criteria would be looking at ALT for defining drug-induced liver injury. So, Leonard?

Seeff photo, biosketch, abstract

LS#1: DR. SEEFF: Well, over the past several years in preparation for these meetings, my good friend John Senior would come to me and say, would I please give a talk on the role of the ALT as a quote "biomarker" for drug-induced liver injury. And like a good soldier, I respected the invitation and say, yes, I would do so.

And I have done this several times. Each time it has been exactly the same as it had been the year before. So, it has been a very boring presentation. So, he came to me this time and said, "Well, you must have learned something. Maybe you're going to make a better job of it this time."

And I regret to say that what you will hear is exactly what I have been saying for the last several meetings. So, please sit back and be prepared to be bored by all of this. (Laughter.)

LS#2: As we know, at the present time there is no definitive biomarker for drug-induced liver injury. The whole issue, then, of causality assessment of DILI is an inexact science.

We are still dependent on serum enzymes, particularly ALT, but also the AST and alkaline phosphatase, as a trigger of interest. It is not a diagnosis of drug-induced liver injury, but it is a trigger to tell us that there is some liver injury.

And so, the problem we have is: what are the levels of ALT abnormality that we should consider to be of importance? There are a number of shortcomings, as you know, with respect to the ALT as a, in quotes, "biomarker". I list four here.

LS#3: The first is that it is a non-specific marker of liver injury and not of drug-induced liver injury. It requires, then, as we have heard,

that all other causes of liver injury first be excluded before you can consider the possibility of incident drug-induced liver injury.

Second of all, there is no consensus on what level of ALT increase during the treatment should signal possible incipient drug-induced liver injury. Should it be three times? Should it be five times? Should it be eight times?

Thirdly, as we have heard this morning, whatever the agreed-upon ALT level, it has to be measured against a background, a comparator, and the comparator is usually the baseline level. The baseline level that people consider to be important is what we call the, in quotes "upper limit of normal." As we have heard today, there is a lot of dispute on what the upper limit of normal is.

LS#4: Finally, there is a whole new paradigm, and that is, now that we have the very potent antiviral drugs for hepatitis C, we are bringing into the fold people who already have abnormal enzymes to be treated. And so, how do we then use that same abnormal enzyme as a biomarker, as a

marker of potential DILI? So, these are the four issues that seem to be we have to consider in the whole issue of the ALT as a biomarker.

LS#5: It is also complicated by the fact that there is a complete difference between assessing drug-induced liver injury in clinical trials as compared to assessing it in clinical practice. In clinical trials, mostly people who come in, the choice generally, other than people who have been treated for chronic liver disease, is to bring in people who start with normal enzymes, below whatever level is considered to be the upper limit of normal.

Second of all, these individuals are generally monitored for the possible development of drug-induced liver injury. So that, if we begin to, then, consider what components we have to take into account in trying to assess, look at causality assessment of drug-induced liver injury, we have the components.

In clinical practice, on the other hand, we learn about drug-induced liver injury when a

patient has been on a particular drug for a while. They develop symptoms. There has to be something that causes concern to the clinician. They either have symptoms or they have jaundice, and they are on a drug. You think to yourself, well, maybe this is drug-induced liver injury.

And so, in this case we do not have a baseline. We do not have preexisting serum enzymes in which to make comparison. So, the role of the ALT, then, is somewhat different in this setting than it is in the situation of a clinical trial.

LS#5: So, I would like to go through each of these four areas, as I have just mentioned. Let me start off with the impact of the ALT as a non-specific diagnostic marker.

The diagnosis or the consideration, not the diagnosis, the consideration of potential DILI is a circuitous event. We cannot, on the basis of an abnormal ALT, call this drug-induced liver injury. We have to go through the whole process of screening, as Naga mentioned earlier, for the hepatitis viruses, including hepatitis E that

people sometimes forget, for autoimmune hepatitis, for alcohol, et cetera, et cetera, or fatty liver disease.

The result of this is that it is a very protracted assessment that has to take place, which is costly. If we had a specific test that would tell us this is drug-induced liver injury, we wouldn't have to pay for all these tests. So, the downside of having a test which identifies liver disease, but not specifically drug-induced liver injury is cost. At least one part of it is cost.

The other part of it, of course, is that there may be a drug that is withdrawn prematurely while the patient is being evaluated for the possibility that the problem is, in fact, drug-induced liver injury. This may have a negative effect. So, you know, it would be, then, ideal or much better for us to be able to have a very specific test that says this is drug-induced liver injury, which the ALT is not.

LS#6: Secondly, we come to this issue of what should the level of abnormal ALT be that signals

possible impending DILI. Let me again reemphasis -- and I guess Arie has made this point and I think Naga as well -- that finding that normal ALT does not diagnose drug-induced liver injury. It tells you that there is some abnormality, and the important thing to do is to go through the whole process of causality assessment, which largely requires you to exclude all other causes before you come down to the possibility that this is drug-induced liver injury.

Most trials have required biweekly or monthly ALT monitoring to screen for possible developing DILI, which gives us the components in which to assess identification of abnormal levels. But the question is, what should that level be that leads you to worry about the possibility of drug-induced liver injury?

LS@7: It is my view -- and maybe the experts in the field here are going to tell me it is quite wrong -- that, in general, ALT levels that are in the normal range, whatever that happens to be, remain pretty normal over time in persons who don't

have liver disease. So, therefore, I suggest that as little as a twofold increase of the ALT in people who are monitored with prior normal values can at least raise concern for impending DILI in the appropriate setting. That is people who are taking drugs and within six to nine months this abnormality happens.

However, obviously, this level of ALT increase is rarely accepted as a reason to withhold a drug, unless it is accompanied by symptoms of jaundice, although to me it does mean, if I were a clinician, and even if there were two times elevation, I would say perhaps I had better check the blood, the test again in a couple of days and see which direction things are going in.

So, therefore, higher levels of signals for possible DILI have been suggested, but I must state that, in my view, whatever level of cutoff, it is all arbitrary. We don't really have a number that really tells us that this is more meaningful. As we heard, is it 265 or 275? Whatever the choice is, it is somewhat arbitrary, but we need something

to hang our hats on. So, what are the levels that have been suggested? Well, as we have heard, it is either a threefold increase or a fivefold increase or maybe even an eight- to tenfold increase.

LS#8: There is a recent effort to look at this issue, and in an important paper it was suggested that a fivefold increase should be the level that leads you to be concerned. The reason given for that is that you don't want to bring in people who have underlying chronic liver disease that is not easily recognizable, such as fatty liver disease or maybe chronic hepatitis C.

I believe that certainly in the setting of the clinical trials this is not an issue because you bring people in. By and large, you have normal values and you have sequential values that you test and they are normal. People with fatty liver disease are not going to suddenly develop an abnormality. So, I think that fivefold is too high a level, and I believe that a threefold increase should be a more practical and appropriate signal

for possible impending DILI, not a diagnosis of DILI, but to consider the possibility that this may be a problem that needs further evaluation.

LS#9: Well, given the fact that there is a threefold or fivefold, it is against what? What is the comparator baseline level? Well, as we have heard, normally, the traditional baseline comparator is regarding as the upper limit of normal, but this is determined for each lab, as we heard, by screening large local populations believed to be healthy and, then, selecting as a reference number the mean value plus or minus two standard deviations, as we have heard. Unfortunately, the screened local population is not always healthy, some having unrecognized NASH, causing raised ALT levels; thus, increasing the mean level of the ALT.

LS#10: If you have clinical trials and several research labs participate in this trial, it is likely there will be differing levels identified as the upper limit of normal. This will, obviously, have an impact on defining the fold

increase. Now I say this although I also believe, as Arie has suggested, that this is nuance because, whatever level we choose, it is really uncertain whether this is exactly the level we need. But we have to start somewhere.

So, I think to avoid this problem, the baseline comparator used to establish a given ALT fold increase would have to be the upper limit of normal established by each laboratory separately, if there are several labs involved, and not a hypothetical group normal value. A preferable alternative, in my view -- and I have suggested this before, and I don't think that it has taken off -- is that, ideally, I think it will be better to have each individual's own baseline level as the comparator. So, you would, then, start out with a person, have their level, and any subsequent increase of three times is made on the basis of their own baseline level that they started with.

LS#11: Now let's get to this issue of DILI monitoring guidelines for persons with preexisting abnormal ALT levels. As you know, with the advent

of these wonderful, new drugs, the majority of people who have been treated for chronic hepatitis C respond dramatically. Over 90 percent lose their virus and enzymes soon come down.

Remember that chronic hepatitis C is the focus of attention at the moment, but in the future there may be other forms of chronic liver disease, such as fatty liver disease, NASH, NAFLD, and so on, that will be subjected to treatment trials starting with abnormal ALT levels.

For these individuals with already raised ALT levels at the outset, the use of ALT, then, as a signal for a possible DILI is a problem. Of course, the FDA guidelines, as you know, did not establish any rules for persons who enter trials with abnormal ALT levels to begin with

LS#12: So, as I mentioned, the DAAs are highly effective. What happens, within a couple of weeks, the serum enzymes come down to normal; the virus disappears, but they continue to be treated for approximately 12 weeks, although that may change in the future. During that time, the values

may go up again. And then, you have to worry, could this, in fact, be potential DILI? So, this raises the question of what the comparator should be if an elevated ALT develops after successful treatment. Should it be the original baseline level that was abnormal or is it the new on-treatment normal ALT level?

So, for those people who are responders to the DAAs, the Direct-Acting Antivirals, the majority of them return to normal on treatment, and if followed by a rise in the ALT levels, assessment for possible DILI can no longer be measured against their baseline level since the ALT levels may have declined because of treatment. A subsequent increase should, therefore, be compared with a new on-treatment level of ALT if the concern being raised is the increase exceeds two- to threefold. Obviously, reappearance of HCV-RNA would suggest, in fact, that there has been a viral breakthrough and not drug-induced liver injury.

LS#13: What about the very few people who have been treated and coming with chronic hepatitis C,

are being treated and don't respond? This is relatively uncommon. I would suggest that they should be measured against, again, their own baseline abnormal ALT level, not some hypothetical group. In this case, I even would suggest the following approach: if the abnormal baseline ALT value does not exceed 100 units per liter, a threefold increase over the baseline would represent an increase of some concern and lead you to begin to wonder, could this be drug-induced liver injury, and then, to exclude all the other cause.

LS#14: If the baseline level is abnormal, but exceeds 100, I would get concerned if the ALT increase doubles and I wouldn't wait for it to be triple before I would get a little bit concerned about that.

LS#15: And what clinical practice? Well, as I have mentioned, we don't have all the baseline information in clinical practice that would permit us, in fact, to look at an ALT compared to baseline, because we don't have a baseline. This is a

different beast altogether, and this is the whole area of causality assessment that a number of people, some well-known people in this audience have been studying all the time.

LS#16: And causality assessment, then, rarely I think is not based on the level of ALT. In fact, the ALT simply confirms there is liver disease. These are patients who come in who are being studied because they feel lousy, they've got fatigue, they've got muscle pain, blah, blah, blah. They're jaundiced. You do serum enzyme testing and they have an abnormal ALT, and it doesn't matter what the level is, at this point you, then, go through the causality assessment that I guess we are going to hear from Naga at the moment and that DILIN is doing and the other well-known groups in this audience are involved in doing.

LS#17: So, this is exactly what I was trying to say. I am going to conclude with the following: the lack of a definitive biomarker for DILI places the burden on screening for possible emerging DILI on the ALT value. There is presently limited

consensus on how best to apply the ALT for this purpose. Greater consistency in setting the parameters of its applicability is needed until it is replaced by better means of detecting incipient DILI. And this is where all the wonderful efforts that are being undertaken at the moment to identify more specific biomarkers becomes extremely important. And just to remind you that the approach to screening for DILI differs depending on whether it occurs during a clinical trial or in clinical practice. Thank you very much. (Applause.)

Session IB:Moderators, Speakers -2

DR. CHALASANI: Our next speaker is Dr. Ted Guo from the FDA. He is originally from Shanghai, did his PhD at VCU. He is Dr. John Senior developed the eDISH program. Now, also, they are developing an likelihood ratio test (LRT), as described in the biography. He will talk about eDISH.

Guo photo, biosketch, abstract

TG#1: Good morning. My name is Ted Guo. I am a statistician. You might immediately think I am a wrong speaker showing up at the wrong conference. But should I talk? Actually, I talk about the medicine. And actually, I have never attended medical school. I don't know anything about medicine. And if I talk about statistics, you know, this is really the wrong place to talk. So, what should I talk about? Maybe I should just tell a story about eDISH. (Laughter.)

About 10 or 12 years ago nobody knew what eDISH was, so let me tell you a story about how I got to know Dr. Senior. Maybe 12 years ago, we were all in the Parklawn Building. FDA was not on a big campus. It was just in one building and people were very close and nobody locked their office doors. One day when I was working, across the hallway there is a figure entering a vacant office. I didn't know who he was, and he was looking for something because nobody was in the office. The previous employee left the job. And then, there was someone knocking on my door who

said, "I have some problems dealing with the data." And I saw he was Dr. Senior. He was holding four boxes of 3.25” floppy disks, if you remember them, four boxes. (Laughter.) Since the boxes were sealed, never opened, never used, apparently nobody had analyzed the data in the boxes.

So, I just said, "Let me see what I can do. I'll just use my own time to convert all the SAS dataset to an Excel file." Since then, we started to talk to each other, you know, about different things. Talked about our families, where we travel, and we got to know each other. Then, he came one day and said, "You know, I have an idea. Maybe you can do something about it."

If I can go to the next slide, this is the first graph of eDISH, but we didn't call it eDISH then. He talked to me about what DILI is and how to diagnose DILI. He used Excel for the data, and he is very good at Excel. He talked to me, but I was kind of ignorant. I said, you know, what's the big deal? You know, if the patient has a liver problem, do a biopsy of the liver and put tissue

under the microscope; you will see the drugs in the liver and you will see how that affects the liver. So, what's the big deal? Can you do that?

Then, he started to educate me about diagnosis, medical differential diagnosis and what that is. And I just learned. I was very interested in that. We both kind of went out of our own fields, went out of our boxes, and started to talk to each other. And I say, "Well, this is something I can do. You can do that in Excel; I can do that in SAS." Excel is very nice software, but it is not designed to handle a large quantity of data. And I can do that in SAS and I can handle all of the data, 3,000, 4,000, 10,000 subjects. And he then showed me the second graph. He said, "If a subject is located in the upper right quadrant, that may be a potential Hy's Law case. Then, I want to know what happened to that subject." And I said, "Well, maybe I can do something. I can do that in SAS. I can draw that plot. I can do an individual patient time course data record."

Talking about it today, there are now a lot of software-makers. They all say they can eDISH. This is a good thing; it is welcome because, before that, nobody knew what the eDISH is. Now some software-makers and government contractors include eDISH in their software. So, I think we need to look at it closely to understand better what that is. That first graph I showed you, is that eDISH? Is that all? I know some statisticians try to apply some statistical methods to diagnose DILI. Can DILI be diagnosed from the numbers with statistical means?

You know, I am a statistician myself. I know something; you need to diagnose and predict; you need to do a causality analysis assessment. You need some more information, not just a first graph. You need to know the time course data, and then, you need to know something more. You need to read the subject's narrative, something like that.

As we worked together, and I was really educated by John; I learned a lot, even though I don't know anything about medicine. I kind of got

his idea. I think I produced this in SAS, and I knew there is a nice feature. You can make each subject a hyperlink. You can move your mouse over any subject and click it, and it will immediately drill down to a time course data for that subject. I thought that's nice. I know that. Not many software people are using that feature.

But that is not ideal for a tool for everybody to use because you have to install some software on each individual, each user's machine. And it would be nice to have the data and program on a server, so people in my organization can just go online and use it.

Fortunately, SAS has a feature called SAS Intranet that allows you to deliver this on a network. So, I think I kind of had an idea how I could do it, but it took a long time to figure it out, to make it work. And I think for this, actually, John needs to explain it.

Now the third part, the theme of this conference I know is to get it right. Getting it right is to head in the right direction, to use the

right methodology to get the right diagnosis. And the most important thing, first, you need to get the data right. If the data are messy and don't give you the right information, no matter how you do it, you never get the right answer.

So, when we developed this tool called eDISH, working together, we spent some time to think about how to get the right data. We need to tell the sponsor what form of data we should include, what kind of variables we should include. We don't want to include many, many things that are unrelated to this software. So, this software is designed to be simple, to be small, to get the job done. We are not here to compete with other sophisticated software developers. So, we keep this in mind.

Then, we developed a spreadsheet that allowed people to download it, sent it to the sponsor, tell them exactly what we need to do. As a result, after we receive the data, usually in most cases within 30 minutes I can upload that to a server and people can start to run this tool eDISH to examine the data. No instruction, no training

is required, no user manual is required. They just follow the few simple steps, and they can do it.

Now the most difficult thing we need is the clinical narrative. I know that some software developers claim they can make eDISH. I talk to those developers. I said, "How do you handle narratives?" They said, "Well, we generate the narratives automatically, programmatically." I ask, “from where?” They say, “from the case report form, from the sponsor's data.” I thought this is not good -- I talked to John. John said this is not relative; it is not the way a narrative should be prepared.

So, we modified our data requirement. And recently, John spent some time writing a paragraph for who should write the narrative, what it should include, the purpose of the narrative. I think this is a very important piece of the puzzle.

So, we have three steps. First, the eDISH graph one, the purpose is not to predict or make diagnosis; it is to separate out from the mostly

normal subjects a handful of subjects of special interest. That is all its purpose.

And we have dividing lines, two times, three times the upper limit of normal for bilirubin and ALT. And there is a reason to do that. Using statistics I found out that three times and two times, actually, are very close to the 95 percentiles. That makes sense because, then, that indicates a rare event. For a rare event, if you use 95 percentile as a cutoff point without any medical background, it is pretty close. And DILI is usually a rare event.

So, in our current version of eDISH, the threshold or the dividing line is fixed. Now we need to find a way to make it flexible. So now, probably it is time to talk a little about eDISH-2. So, what is that? Is that a second version? Is that an enhancement of eDISH-1? Is that an idea? Is that a necessity? I don't know. Before this meeting, I thought maybe we should talk only about eDISH-1, but I changed my mind in the last minute. I thought the speakers before me were talking about

the change from baseline, talking about the measurement of ALT, there are a lot of issues. As a tool, eDISH should serve a purpose. We should have flexibility to change from the upper limit of normal to looking at a change from baseline. We should be able to change the thresholds. And after we change the threshold, what are we looking at? So, there are a lot of new issues.

So, there are a lot of problems about how to improve the narrative. We still don't have a good solution, but we want to improve that part.

And this, just to show one advantage of our eDISH. I think every time we receive the data, if you run it once, the data are further standardized, installed in one single location, unlike some other software. Data are still either in the users' desktop or still scattered around in a network. But all the eDISH data is on one server. That is a good thing. We can in the future do research, a pool of data. That is very convenient.

Let me see. Now eDISH-2 should have more data than just serum chemistries. We are now

talking about treating patients with hepatitis C or B, we need to get the viral load data. Sponsors submit that. We should include viral load data. If you look at the graph, the lines connecting black dots, that is viral load data. And the lower part is our data requirement. We add the viral load. That is the data on demand. You know, that is when the drug is treating hepatitis C. We probably need to ask the sponsor to submit that.

So, we continue to improve our data requirements, to enhance the tool. The tool is not only a review tool; it should be a review tool and a research tool. So, there is a lot of work to do.

How did I do? John, Are you going to talk about the narratives? Maybe you can start. I think John is going to tell you something very important; that is the narrative. That is something we encountered many, many times. Sometimes we got a lot of narratives generated automatically by machine. It is like a data dump. It doesn't serve any purpose. And a lot of times we found that people don't understand what the narrative should

include. People ask, you know, "We have to generate it by hand, by writing." Yes, true, you cannot generate this by machine. You have to write down what the cause of those abnormalities. So, John is going to talk. I have already spent a lot, too much time on this. Sorry about that. John?

Senior photo, biosketch

DR. SENIOR: That's all right, Ted. Can you all hear me? (Chorus of yeses.) Okay. I don't want to go up on the stage because I am so unsightly. I cut my lip, shaving this morning, and it kept bleeding because I'm anticoagulated on warfarin.

I think there are just a few more slides. I am not going to read them. You can read them faster than I can say them. What I will say is that Ted Guo and I have been working across these two major disciplines of statistics and medicine. Now statisticians are trained and are very skilled at analyzing data, but what they are not trained to do is to diagnose patients, because that is what medical doctors, physicians, do from day one in

medical school. When they see a patient, they immediately think about what is the cause of the patient's problem. That is a diagnosis. Why is it important? Because medical doctors, as distinct from other kinds of experts --- pharmacologists, toxicologists, chemists, and statisticians, --- medical doctors have a responsibility to treat patients, to write prescriptions, to order tests, to do something to treat the patient that nobody else has the responsibility or authority to do. And therefore, if they are going to treat the patient correctly, they have to make the correct diagnosis. So, they are the only ones who make diagnoses.

Trying to get statisticians to understand the process of medical differential diagnosis, finding the cause, is not so easy. And eDISH is a diagnostic tool really to help the medical doctor make a diagnosis using statistical data.

For eDISH we are using the computer and we are using the human mind. They are different. The human mind is very good at recognizing patterns,

recognizing faces. I can recognize Ted Guo. I can tell Arthur Karmen in a half-second, even though I haven't seen him in years. I can immediately look at them and in one glance I say, "That's Arthur Karmen," "That's Patrick Kirby," "That's Naga Chalasani," "That's Leonard Seeff." I can do that zip, zip, zip. A computer can't do that. A computer can't recognize faces as well as humans can. We have been trained over eons of time to protect ourselves by quickly recognizing friends and enemies. So, we recognize our friends and we greet them. The enemies we stay away from.

So, computers and human minds work in different dimensions. On one hand, as Ted Guo said, I came to him with disks, with data from thousands of patients entered on them. It would take me weeks to do what his computer can do in less than a second, analyze the data. The computer is powerful in quickly analyzing data that the human mind cannot deal with.

TG/JS#13: On the other hand, once you display the data on all the patients and ask “which ones have

both elevated bilirubin and ALT?” and you display that on a graph, a human mind can look at that graph in a blink of an eye and say, "Oh, I'm interested in the few patients up in the upper right quadrant. I don't want to worry about the patients whose data are all normal." So, I can recognize the pattern at a glance.

But, then, we say, let's go back and have the computer tell us everything that is in the data about that one person. Give us a time course. Now a time course is really adding a third dimension. One dimension is the ALT; that is the injury dimension. Another dimension is the bilirubin. That is the dysfunctional measure, a second dimension. And the third is time. How do the data change over time, day by day, week by week? Is the ALT rising? When does the bilirubin rise? Is it falling? That is a very important piece of information that helps diagnose what is going on.

So, putting the time course up, which takes the computer another half a second, allows the physician to look at the time course and interpret

how the values are changing and what that might mean for making a diagnosis.

TG/JS#14: But we then need a third step. We need additional information that was not in the protocol. It was not in the case reports. And there is no use searching in the case report because the information isn't there. It wasn't anticipated as needed when the protocol was originally written. You never can anticipate all of the things that might happen.

So, the protocol and the case report cannot or should not be used to create a narrative. You have to go back to the physician at the site, the investigator, who is a medical doctor. All of the investigators of clinical trials are MDs, for a good reason, because they have responsibility to protect the safety of the subjects participating in the study.

TG/JS#15: If you are experimenting with a new drug, there might be injury. It is important that the physician in charge of the study site be ready to interpret what is going on and, if necessary,

stop the drug or interrupt treatment with the drug, or do something else, maybe take further action. But certainly, diagnostic consequences follow what the data show at the study site.

So, Hy's Law is not diagnosable by just chemistries. You need more information. You need the time course, not just one day, but the changes over time, and supplementary information. You need to know what other information the physician in charge at the study site used to determine whether or not the drug should be stopped or not. It is not in the chemistries; you can't do it automatically by chemistries alone.

So, preparing the narratives should be done thoughtfully by the person who is responsible for the welfare and safety of the study subjects, not by a data analyst, not by a clerk, not by a project manager, but by a physician who has the legal responsibility for the safety of the subject.

TG/JS#16 Clinical trials are more than just gathering papers and numbers to get an approval. The safety aspect of clinical trials is very

important, particularly in the field of liver injury, because any drug is potentially capable of injuring the liver, sometimes fatally. In 1997, the FDA approved eight drugs -- eight -- that had to be removed from the market because they were killing people. Four of those were for liver problems, three for cardiac problems, one for muscle problems. But four out of the eight were liver toxicity. That was the year that triggered off the first of these conferences. I will say, as a result of that, the FDA has raised the consciousness of its reviewers, and the reviewers have raised the consciousness of the pharmaceutical companies. As a result of that interaction between the reviewers at the FDA and the sponsors making the new drugs, no drug has been approved by the FDA since 1997 that has had to be removed from the market for fatal liver toxicity.

It is not that minor toxicity doesn't occur. Sure, we get transaminase elevations. So what? Mild injury is not killing the patient until the function of the liver is so badly disturbed that

it can no longer do its job. We can cut out two-thirds of the liver and throw it in the bucket, or we can injure chemically two-thirds of it so that it is not working. And the liver is still able to regenerate, remarkably, more than any other organ that we know of, and the person lives.

So, all of this has to be factored into what Leonard was just talking about, how you make the right diagnosis of what caused the problem. How do you know that what is happening was caused by the drug and not by some disease, not by something else? That is not so easy. It must be done by somebody who has spent a whole lifetime making diagnoses, and that is a physician.

We are now going to hear from Naga Chalasani, who has done some extremely interesting work.

Chalasani photo, links to biosketch, abstract

NC#1: DR. CHALASANI: Thank you, John. Nice commentary there. I don't have a formal disclaimer conflict slide that we do for CME, but I do want to disclose I think I am a boring speaker, I think.

The American College of Gastroenterology commissioned this practice guideline in 2012. It took about two years to write. It was published in June 2014 in American Journal of Gastroenterology. Most of the authors are from the drug-induced liver injury network (DILIN) which, as many of you know, is the group chaired by Paul Watkins, and Jose Serrano is the Program Officer from at NIH.

It is an interesting story. There have been stories all morning, so I will tell you a story as well. I belong to the group that identifies topics and assigns authors for these Practice Guidelines. One morning the meeting was in Vegas. I was late. Too much poker the night before. This guideline writing was assigned to me. (Laughter.)

NC#2: Some of you may have worked on practice guidelines. They are not easy to write; they need a lot of consensus-building, compromises, and the society leaders have their own opinions, et cetera. So, it is not an easy thing to do. Skip Hayashi was the second author on the practice guideline, and

he is here, with Vic Navarro, Will Lee, and Bob Fontana, who also have done wonderful work.

NC#3: Practice guidelines use different elements to attribute the strength of the recommendation as well as the quality of evidence. As you will see in this guideline, the quality of the evidence ranged anywhere from low to very low. Basically, the editors asked, "Show us the papers. If you cannot show us the papers supporting what you are saying, it doesn't matter how strongly you feel. It just has to be low or very low evidence." That is what it came down to. And the strength of the recommendation is clinically how strong they feel, whether it is a strong recommendation or a conditional recommendation.

So, the practice guideline had a number of summary statements and recommendations made that I will review. These are about 16 of them. Let me just walk through. Some are pretty straightforward. I would like highlight some that seem very strong, common-sense recommendations, a no-brainer, and yet, you will see a very low level

of evidence, just simply because there are no published data supporting that.

NC#4: The first recommendation was about patients with suspected hepatocellular or mixed DILI. Acute viral hepatitis and autoimmune hepatitis should be excluded with standard serologies. I think what made it to the practice guideline is hepatitis C RNA testing. In some industry case report forms, you see hep C RNA, but in the clinical setting it is not being done. They just do an antibody, which can be falsely negative. And in the DILIN prospective study we incorporated hep C RNA, but I it has not been done on a consistent basis.

There was a lot of debate about anti-hep E IgM testing. I think in the DILIN prospective study there were about seven cases early on that Tim Davern was the first author in the gastro paper where, for all purposes, the cases looked like DILI. When you dug deeper, when Bob Purcell at NIH did the serologies, there were about seven cases that were potentially acute hepatitis E. But one

of the reasons that did not make it to the practice guideline is there is not a commercially-available test. Once again, keep this in mind. This is for practicing clinicians, not for researchers. So, here we could not recommend routine anti-hep E IgM testing, just for the lack of -- there is no standardized testing. But, as you work up in the drug development Phase 2/Phase 3, if you see it, I think it is important to consider hep E IgM.

Whereas, acute CMV, acute EBV, acute herpes simplex are based on clinical scenarios, if there are, for example, lymphadenopathy, atypical lymphocytes, then you test for these. Finally, Wilson's disease and Budd-Chiari are also on a case-by-case basis rather than on all-comers. And again, really low to very low level of evidence.

Imaging did not make it. It is done a lot in the clinic. Just about everybody with DILI gets a liver ultrasound or a CT, even for AST like 9500, generally speaking, overused. But in the practice guideline for hepatocellular and mixed DILI there

is no imaging required, whatever it is worth for industry investigators here.

NC#5: For suspected cholestatic DILI, I think abdominal imaging is pretty straightforward to exclude DILI pathology. And I cannot tell you how important it is. Sometimes you may even want to repeat it few weeks later, especially if alkaline phosphatase is continuing to go up or bilirubin, because I think early on we did not see dilated bile ducts two weeks later that showed up, but there were cases of pancreatic cancer picked in DILIN prospective study where we enrolled based on early ultrasound that was negative.Serological testing for PBC should be limited to selected cases. And MRCP or MRI -- excuse me -- MRCP should be quite limited as well, although in clinical practice this is, once again, quite more utilized.

NC#6: When to consider a liver biopsy? Dr. Kleiner is sitting in the audience; he has published really wonderful papers recently in this area. Biopsy is optional. So, a lot of low level of evidence, which just generally means this is a

lot of consensus and expert opinions rather than published studies. A liver biopsy should be considered if you cannot exclude autoimmune hepatitis. I don't think anyone would disagree, common sense, but, not one rigorously tested. So, I think this was a low level of evidence, especially if you consider immunosuppressive therapy. Liver biopsy may be considered for a number of sub-bullets here. There is an unrelenting rise in liver bile chemistries or signs of worsening liver function, despite stopping suspected offending agent. I think the reason behind this is twofold. One is to see amount of liver damage, whether there is necrosis or the amount of necrosis, and, also, to see if there is unsuspected autoimmune hepatitis or other pathology. For example, we have picked up some cases of ischemic hepatitis, which seemed like drug-induced, but undetected right heart failure and arrhythmias, et cetera. So, that was the basis for that.

And after stopping the compound, if the DILI is not resolving as you would like to -- for

example, if the ALT has not fallen by 50 percent within a couple of months or, for example, peak alkaline phosphatase has not fallen by 50 percent at six months -- to see if the patient is evolving into some form of chronic injury, whether chronic hepatitis or vanishing bile duct syndrome. Especially in chemotherapeutic agents, if you need to re-expose the patient to the same compound. It also happens to some degree in the IBD area. You need to give the same biologic agent because the patient needs it.

Then, it is a consideration. This more happens with the low levels. A patient may have underlying NASH. In clinical settings sometimes you may not have baseline. If you started a biological agent and you have an ALT of 90 or 100, you don't know if it is a new onset, whether you want to stop. Is it underlying NASH. That is another reason. Obviously, if liver test elevations, after a DILI episode if they are not resolving at 180 days -- or, actually, I think it may be Bob Fontana's paper will say, after the onset

if you have persistent abnormalities, consider a follow-up liver biopsy or consider a biopsy to see if the patient has evolved into chronic DILI.

NC#7: Bullet 4, the recommendation that re-exposure is strongly discouraged. This was discussed at many of the meetings here. I think Chris Hunt had some data from Glaxo at the time. Especially if the initial episode was significant -- I am not using the word "clinically significant" -- but if the ALT were five times or Hy's Law, et cetera. Of course, in medicine you cannot always meet case by case. There may be an exception made there is not a suitable alternative.

Recommendation 5, in individuals with suspected DILI, especially when liver biochemistries are rising rapidly or there is evidence of liver function, suspected agents should be stopped promptly. Does anyone disagree? It is a strong recommendation, super-strong, but the evidence is low level.

The sixth is about no definite of three treatments. NAC could not be approved. There was

some soft data with acute liver failure, drug-induced in children from Will Lee and Bob Squires. It said not recommended in children, but in adults it is a soft conditional recommendation that drug-induced acute liver failure, NAC could be used because in the NAC trial there was a subgroup that showed benefit, at least a trend towards benefit. So, that was a conditional recommendation with a low level of confidence.

NC#8: Recommendation 8 was about herbal and dietary supoplements (HDS). We are seeing a lot and this is getting a lot of publicity. Patients should be encouraged – (directed at clinicians and medical monitors in clinical trials) -- to report the use of HDS, and they should be reminded that the supplements are not scrutinized at the same level or not at all in some instances as drugs. And the diagnostic approach for HDS is sort of evolving. Especially, the difficulty is multiple compounds are taken at the same time. So, it is not easy to attribute to a single compounds. You don't know the signatures. Nonetheless, I think

the exclusion, the severity, causality, adjudication issue just generally followed the same guidelines.

Recommendation 10 is straightforward common sense, which is, if you suspect HDS-related DILI, stop the compounds. Once again, low level of evidence.

Recommendation 11 is about DILI in patients with chronic liver disease. Underlying chronic liver disease requires a high index of suspicion. There is a paper from DILIN that is on the MedLine, an early ePub, that describes about a paper that had 900 DILI cases, and about 90 of them have underlying chronic liver disease. It seems like a majority were fatty liver disease. And there is a two-by-two comparison of DILI in patients with and without chronic liver disease. I think if you get a chance, you may want to look at some of those.

But the point is that it is extremely difficult, especially with patients with hep B, hep C, when they have an increase in ALT or bilirubin, to know if the cause is underlying liver disease,

as opposed to what is drug-induced. Keeping high vigilance is desirable.

When you look at the package inserts, especially for many of the compounds, there is frequent biochemical monitoring in patients with underlying liver disease. When you look carefully -- I know Janssen has written a nice commentary on this as well -- really you don't find much data supporting that the careful biochemical monitoring prevents, actually, when clinicians follow. In the case of statins, everybody knows less than 50 percent of the providers prescribing statins do the tests in a fashion that was at least early on in the package inserts.

NC#9: This recommendation reads, "There are no data to recommend specific liver biochemistry monitoring plan when a potentially better toxic agent is prescribed in patients with known chronic liver disease." This is a big problem, though. Especially if you think about fatty liver as a known chronic liver disease, you are talking a third of U.S. adults. Often the information contained in the

package inserts is incomplete or unhelpful. Once again, I think Einar's commentary in Gastro two or three months ago is quite instructive. For the same compound by two different manufacturers, it might have entirely different recommendations. For example, sumatriptan made by one company would have one instruction, one warning; whereas, made by some other company would be totally different.

Patients should be advised to promptly report any new onset symptoms such as yellowing of their eyes, abdominal discomfort, itching, dark urine. That is pretty straightforward.

We thought it was reasonable to monitor liver biochemistry at four- to six-week intervals, especially during the initial six months of treatment with potentially hepatotoxic agent. A very soft recommendation. This is more so on consensus rather than evidence. That is how this received a Michelin one star. (Laughter.) A really low level of evidence.

So, practice guidelines rest on compromise, expert consensus. Especially in a field like DILI, there are not a lot of randomized trials. I am going to stop there. Thank you. (Applause.)

Session IB:Moderators, Speakers -4

DR. CHALASANI: Our last speaker this morning is Dr. Patrick Kirby, who will talk about normal ranges for 12 novel biomarkers for liver safety. This is an interesting dataset.

Kirby photo, links to biosketch, abstract

PK#1: Thank you. I appreciate the opportunity to come here and talk today. I am a toxicologist in non-clinical safety. I work with rats, dogs, and monkeys, so I think about things in slightly different ways than human doctors. I am part of the Predictive Safety Testing Consortium (PSTC) funded by the Critical Path Institute, working on discovery of new drug development tools in certain gap areas. And one of those is liver injury. I have only one ALT slide to show you today. The whole

idea behind this talk is whether we can find other tools to help where ALT is not giving you the information you want. This particular project is being done in collaboration with SAFE-T (Safer and Faster Evidence-Based Translation). They are a group funded by the IMI, looking at new clinical biomarkers. We are specifically looking at drug-induced liver injury. The leaders of the SAFE-T group are here today, Michael Merz and Gerd Ublick, and they can address any specific questions on some of their goals.

I am just going to talk to you about baseline values. I am not going to show you thresholds for concern, such as 3X upper limit of normal. I am just going to show you the baseline values for 12 biomarkers in a normal, healthy volunteer population. Thresholds for concern, different baseline values for different populations, are all to be figured out. But it is a good place to start for some biomarkers that may have novel contexts of use that may be validated by ongoing clinical trials which SAFE-T is running, to understand the

performance of these biomarkers. That is it in a nutshell.

PK#2: The key messages are: We are in a collaboration between the PSTC and SAFE-T. We looked at 12 novel liver biomarkers. Some of them are already in the literature. Some of them already have kits available. Some had to have the kits created by the natural and medical NMI and Tubingen, which is an assay development group. They developed a lot of these assays for us, and I will talk about them. We looked at the normal healthy volunteer (NHV) population, where there was low intra- and inter-subject variability. I will show you that data. We did look at stratification factors. We didn't see anything except for age, and that was only observed in alpha-fetoprotein and prothrombin.

What we use this information for is, hopefully, for eventual clinical qualification of biomarkers that perform well in specific context-of-use areas. A lot of this data is coming out sometime in June. SAFE-T is doing a lot of

clinical trials looking at these particular biomarkers.

PK#3: The background on this was a PSTC study that they had run in the past to look at renal biomarkers. It has already been used to help them set some thresholds and think about what to do in the renal space. The subjects were recruited in Kalamazoo, Michigan at the Jasper Clinic. So, a total of 81 volunteers over three site visits. We collected plasma, serum, urine, and blood on day 1, 6, and 20. These are all fasted samples.

PK#4: This is my only ALT slide. Standard inclusion/exclusion criteria. We looked for hepatitis and things like that. The original intent of this healthy volunteer study was to look at renal function. So, there was an emphasis on glomerular filtration rate. Most of the ALT values were below the upper limit of normal. We had maybe five patients that were above the upper limit of normal. When you look at ALP, we had maybe one patient above the upper limit of normal. For this analysis, we included everybody. In the future,

we could do more sophisticated statistical analysis and pull out subsets. When we did some initial correlation analysis to ALT, everything was all over the place, because I think everybody is within that normal reference range. Those are our 81 subjects. You can see they are all on the bottom. About 5 people were above the upper limit of normal.

PK#5: Demographics. These are people from Michigan, about 81, 41 between 20 and 39, the other 40 between 40 and 70. One thing I would point out, it was mostly a Caucasian population and in addition, these are not light people, with a pretty healthy BMI. Some people were obese, about 33 percent of the population and 60 percent overweight. That is something else we have to take into account.

PK#6: These are the 12 biomarkers that we assessed, listed here. Cytokeratin full length, the caspase cleaved, GLDH, GSTL-alpha, alpha feto-protein, macrophage colony-stimulating factor 1 receptor, arginase-1, osteopontin, PON1,

LECT2, and SDH. These were either serum or plasma biomarkers. The detection was either by ELISA or colorimetric assay, some commercially-available, some developed by NMI, which is a SAFE-T affiliate.

PK#7: I have included a table which has all the information on these 12 biomarkers. I am not going to go through each one, because it would take a lot of time. One thing I would point out, is that Brett Howell of the Hamner is going to talk about cytokeratin tomorrow. So, he will give you some background on that. But the full length is a biomarker of necrosis. The caspase cleaved cytokeratin is a biomarker of apoptosis. GLDH, it is thought to have better specificity than ALT. Maybe you can use it to differentiate ALT sample from liver of muscle. You can look at these. There're references for you. I have this next slide here. I would add that these are a subset of some of the biomarkers SAFE-T is looking at. I know they are also looking at miR-122 and some bile acids. We didn't run those because the assays

weren't ready when we pulled the trigger on these samples.

PK#8: In the Predictive Safety Testing Consortium, we do see some overlap in these assays that we are looking at in rats, specifically, GST-alpha, Arg1, GLDH, miR-122. And so, we could couple this data together and just see how these biomarkers translate across specifies in the various contexts of use that we will discuss. These samples were three years old, in the freezer, at minus 80. For long-term stability, there haven't been dedicated studies done yet for all the various kits, but what was done is that new kits or new lots arrived. They assessed previously measured samples. The data were pretty tight at this point, but we still need to run long-term stability experiments, and that will be done.

For the concentrations we observed, they were consistent with what was in the literature -- so that was good -- except for the CSF and PON1 where there were slight differences with the newly-created kits, but the guys at NMI are

planning to use mass spec to add a correction factor to have more realistic values.

PK#9: The statistical analysis is just for your reference. Basically, we have three types of data, one type where everything was above our limit of quantification. So, it was a little bit easier to do our stats on that. We had a second set that included the caspase cleaved 18 and GST-alpha, where some of the values were below our limit of quantification.

For the full-length cytokeratin, where most of your values from healthy volunteers are below the limit of quantification, because you don't expect that biomarker to be up in normal healthy volunteer patients, we had to use different statistical analysis for different sets of biomarkers.

PK#10: So, this is the real money slide with all the data and information that may be useful, just with all the different biomarkers here listed, you know, the united, the estimated geometric mean, intra-subject CV, inter-subject CV, and then, the

estimated upper limit of normal by the 95th percentile. I would have to say that the intra-subject CV is over the three different site visits. Overall, there was low variation but intra-subject CV was a little bit higher than inter-subject, except for LECT2.

PK#11: And now I can go into some of the raw data, just to give you a view. For the LLoQ, just what samples were below the limit of quantification. For the most part, 7 of the 12 biomarkers were all above the LLoQ. Full-length cytokeratin 18, most of them, 93 percent were below the LLoQ. And so, that is why on the previous slide we don't have this geometric mean, just because you couldn't calculate it. A few others there that were below the LLoQ.

PK#12: This is cytokeratin 18, the caspase cleaved, a biomarker of apoptosis. The red line is your estimated upper limit of normal by the 95th percentile. This is your geometric mean, the lower black line. Then, the three site visits, visit 1, visit 3 and 4. The reason why there are

different numbers is that they didn't collect blood or serum or plasma on visit 2.

PK#13: To give you kind of an idea how patients looked over the site visits, what you are seeing here is all 81 subjects. On the top is SDH; on the bottom is LECT2. The different colored dots are the spread over the three different site visits. So, overall, it was pretty tight for both of them. We were pleased with the data.

PK#14: In terms of stratification factors, we looked at the 12 biomarkers. The first thing I wanted to look at was BMI, just because these were very obese patients, you know, 60 percent of them. We didn't see any stratification based on BMI. We didn't see any stratification based on gender, ethnicity. The only thing we saw was age-related stratification just in two biomarkers, AFP here, where you can see that there is a slight elevation in older patients, greater than 35, ---

PK#15: --- and prothrombin as well. Overall, there was nothing that really jumped out to us.

PK#16: Some of the next steps that we have to do are to look at some additional biomarkers in this patient population because we have a lot of samples saved. That includes miR-122 analysis, potential bile acid quantification, as well as individual bile acids because it is a nice study population.

In addition, as I mentioned, the work package 3 of SAFE-T is in the process right now of looking at all these biomarkers with the same assays in various clinical trials that actually cause liver injury. They are looking at three different context-of-use areas. This kind of gets back to what Arie was talking about before, getting these new drug development tools, if feasible.

PK#17: So, just to show you some of the things they are thinking about, and they will use this baseline data, this normal healthy volunteer data, to help them think about the changes they will see in their different indications in these clinical trials that SAFE-T is running. One context of use that they are looking at is something to confirm DILI beyond just ALT or bilirubin, like is this

sensitive for that? That is a very general context of use. Another thing they want to see, are you going to progress to DILI? Can you use some of these biomarkers to say, hey, this person has ALT, bilirubin, but you also have this biomarker, and you have a greater chance to progression? Can they see that in their patient population?

Then, additionally, can these biomarkers help you detect subclinical DILI where ALT is less than threefold upper limit of normal? The same thing with bilirubin.

So, hopefully, some of these biomarkers will perform. I think when you are in biomarker discovery and you get this big list and you keep on whittling down and down, but the hope is to get new tools that you can really add to your toolbox that can give you information about potential progression of the DILI or predicting which patients will progress to DILI. And so, that is kind of the direction we are going.

PK#18: Just to recap, it was collaboration between two different biomarker consortia, which

is a great activity because so often in the drug development business you are very competitive, but I think these consortias really push -- we shouldn't be competing on safety. This is where we should be really collaborating and working together.

And so, we have some baseline data from healthy volunteers, but with a slightly-elevated BMI. The data look good, some minor age stratification. Hopefully, we will be able to use this data once we see how these biomarkers perform in clinical trials with actual liver injury.

PK#19: To thank people, the Biotoxicity Working Group, the SAFE-T group, a lot of people here, like Will Proctor, Phil Schrott from PSDC, and John Marciniak from Takeda helped me look at the ALT values of a lot of these samples. So, thank you. (Applause.)

Session IB Discussion

DR. CHALASANI: I think we have saved enough time for Q&A. I do have a question for you, Patrick. About the narrative, as part of this adjudication, DILIN has used them for now close to 14 or 15 years, and we have gone through different iterations. No question it is useful, but there is a lot of bias, right? The PIs who send cases into DILIN are trying to sell the cases. We would the stuff like, you know, this is a great case. And then, Jay Hoofnagle would just tear it apart, et cetera. So, I do think if you go the route of clinical, I agree the investigatorss who are taking care of the patients should write narratives, not the medical monitors, because there is a knowledge and perspective gap there. I think having something structured is going to be very important, rather than introducing a systematic bias. That is one comment. I am curious what you think, John.

DR. SENIOR: Well, I think that the quest for new biomarkers is useful, but there is a mistake that is being made which runs across it. They are

all being compared to ALT alone. I think we are now way beyond that. We don't use ALT alone as a biomarker of serious liver injury; we are using the combined bilirubin and ALT. So, it is really an unfair comparison to use ALT alone. It just makes the new biomarker look better than it is. What we really need is a biomarker that is specific for liver, and not only for liver, but for drug-induced liver injury. There are lots of kinds of liver injury. And even these biomarkers are not specific for any kind of liver injury. So, I am not persuaded that any of these new biomarkers is really an advancement of any significance, unless it can be shown clearly by a large margin, better than what we are using right now, which is the combination of bilirubin and ALT. When there is enough injury that the functioning of the liver is disturbed, then it is clinically important.

Now, Patrick, maybe you can comment on whether you are looking at pairs or combinations of these biomarkers, not just one by itself.

DR. KIRBY: Well, I agree with you, that ALT is very sensitive. It is a great biomarker. Some of the contexts of use where ALT may fall down is specificity. Can you use a biomarker to say GLDH? So, your ALT is up due to a non-liver source. I think there are a lot of different things to look at. I think currently, right now, they are comparing to ALT alone in terms of sensitivity. It is very difficult to beat ALT on sensitivity. In terms of a panel approach, it has been discussed, but right now it is just comparing to ALT at this point.

DR. CHALASANI: Before I go to you, Arie, if I understood your slides correctly, the cleaved cytokeratin 18 -- and by the way, we should only call it keratin 18, the new nomenclature, not cK18, just K18 -- the 95th percentile is all the way up to 260 units per liter? I think you are studying fatty liver disease.

DR. KIRBY: Yes, BMI is not low in these patients.

DR. CHALASANI: Yes, but I think there are a number of data coming along. I think using only

180, it is not healthy. I think you are dealing with fatty liver disease. I would suggest that you proteinize your cleaved K18 fragments to ALT. I think you will pick up a lot of abnormal ALT beyond the product criteria.

DR. KIRBY: Yes, we definitely tried to see if ALT, you know, the highest ALTs had the highest for all the 12 biomarkers. It just wasn't the case when we did an initial mapping and plotting it out. They just didn't fall out.

DR. CHALASANI: I would be surprised, though. K18 fragments track ALT so well. I mean, there are just not one. There are now MRI-based PDFFs, so on and so forth. So, I would ask you again to look carefully.

DR. KIRBY: Yes.

DR. CHALASANI: Arie, Sorry.

DR. REGEV: That was my question exactly. Relevant to the K18 and the effect that you find, if I understood correctly, you had both an obese population ---- and it never reached the lower

level of quantification. That is very interesting.

DR. KIRBY: For the full-length.

DR. CHALASANI: For the full-length.

DR. KIRBY: Not the caspase cleaved.

DR. KIRBY: Yes. We did both.

DR. REGEV: Okay. Because, again, this is probably one of the most popular groups of NASH, and that is with patients.

But I had a comment. I was very happy to hear that you did mention a few of the questions that we are struggling with. I think the approach to these biomarkers should first be, what questions are we trying to answer? Which ones are the questions that we are struggling with those biomarkers? Because I keep hearing that we are trying to find something more sensitive than ALT.

DR. KIRBY: No, no, that's not what I am saying.

DR. KIRBY: I am hoping as sensitive as ALT and, hopefully, more specific in some cases.

DR. REGEV: Exactly. Basically, ALT has its function. We need something that will tell us, when ALT is elevated, if this measn this particular patient will end up with liver failure or will he end up by adaptating? Those are the kind of questions we hope those biomarkers will answer. I am sure you are working with clinicians on those questions.

DR. KIRBY: No, I agree with you.

DR. PRATI: I have a question for Dr. Chalasani. Among the different diagnosess that you have included, among those that had to be excluded for the diagnosis of DILI or such as hepatitis B, hepatitis C, and so on, you did not mention acute alcoholic hepatitis that is still at least one of the most important causes of acute liver failure. What is the reason for that?

DR. CHALASANI: I think it is just an assumption I am thinking -- and I can't tell how we -- it is a pretty obvious diagnosis perhaps. You know, that is probably why we did not specifically mention it, but, yes, absolutely.

Not only that, for example, ischemic hepatitis; everything else needs to be ruled out. Although these were the recommendations, when you look at the description of the Practice Guideline, there is a nice differential diagnosis of all the things that one should consider.

DR. PRATI: Because in several situations, especially in Europe, the pattern of, the behavior of alcohol drinking changes. So, some people drink, for example, with binge drinking around the weekend. So, this can cause actually really sometimes acute intoxication hepatitis even in people who are not drinkers.

DR. CHALASANI: Yes, the point is well-taken. It did not make it to a recommendation, but it is described in detail in the narrative part and the summary statement. I think all causes of liver disease, especially acute forms, should be considered and excluded.

DR. WATKINS: Yes, a great session. I want to comment about the biomarkers that Patrick Kirby was talking about and others coming out of the

SAFE-T consortium. What is really happening is a huge opportunity where they are developing high throughput assays for getting the normal ranges. These can be performed in 15 microliters of serum or less. It is not just an issue of comparing them to ALT. So, DILIN Network has a collaboration with SAFE-T -- I see Michael there -- where we have given serum samples from 166 individuals during the acute liver injury, within two weeks of the initial discovery.

And so, getting the samples there, we are looking at the ability of these new biomarkers to predict duration of injury, outcome. About 10 percent of these individuals have transplant or liver failure. So, it is not just an issue of comparing to liver chemistries.

But I think we know enough now to know that this is going to be a different road of trying to really find out the value of these biomarkers. Some of the ones we thought were most promising are actually showing up as being positive in what we know are benign ALT elevations.

So, the road to really understand and find how to implement these in drug development is going to be long and require thousands of patients with different diseases. And that is really one of the central issues in the discussion tonight at eight o'clock. It is, should we be thinking of a broad pre-competitive way to prepare us all in the industry to be able to utilize and understand these biomarkers?

DR. CHALASANI: Once again, I still am very confused. What are the things that need to be addressed? I mean, I think this cross-sectional comparison -- I mean, who with elevated ALT is going to develop acute liver failure? I think we have bilirubin. We have INR. We don't need this new --

DR. WATKINS: Why wait for the bilirubin? We have seen in the example that is given sometimes drugs are stopped before the bilirubin even rises, and it progresses to a Hy's Law case.

DR. CHALASANI: I can see some, but I think that one of the first discussions would have to be clearly finding what is the unmet need that you are

planning to address. Do we need new markers that predict chronic DILI at month 12 or vanishing bile duct syndrome? I don't think we are seeing clearly what are the issues that we want to address.

DR. WATKINS: Well, no, that is even all the more reason to really get the best minds together. And SAFE-T is actually moving right along with developing a context-of-use -- and, Michael, I might say too much -- context-of-use statements to run by regulatory agencies. That is ongoing for the next month. So, I think that it is very important that we make sure that these new biomarkers are interpreted correctly and the limitations are identified.

DR. HONG: I am James Hong from China and representing a medical consulting company in China. We have been working with a group of Chinese investigators for about two years. Since DILI is a serious problem in China, we have been developing a platform where we should go ahead and test ALT. Well, the purpose is to establish a patient

register first. Then, based on this platform, investigators can start to do clinical research.

But, since China is very, very large country, even clinical research can involve hundreds or thousands of patient subjects. However, actually considering DILI is a serious problem, there might be tens of thousands, even millions of patients in China.

So, we have been thinking of deriving a system like eDISH. However, we are considering a large population. So, we have a quick question for Dr. John Senior. Do you have any thoughts of expanding the use of eDISH from clinical research to routine clinical practice? Then, the patient can use the system for self-diagnosis, even alert of some suspected cases. That is my question. Thank you.

DR. SENIOR: Maybe Dr. Guo can answer. He was just over and spent a month in Shanghai, I think in September. He was talking directly to the Chinese companies. I don't know who he was addressing in Shanghai. But, clearly, we have to

be aware of the fact that China has -- what is it? -- 1.3 billion people?

DR. HONG: Right, yes.

DR. SENIOR: A gigantic problem. So, I think that we all need to be aware of this as a global problem. It is not just a local problem. It is all over the world, and we really need to be -- I think probably what Bob Dufour said this morning, that clinicians have to be the ones to put the pressure on the companies, the manufacturers, because that is what worked before. It was when the reviewers put the pressure on the pharmaceutical companies that we got the companies. It was not by congressional law. It was not done by the Commissioner issuing a proclamation. It was done by persuasion from the bottom up.

Now the physicians use ALT. Every one of us uses ALT all the time to detect liver injury. But we are using it in different ways, and we don't have a reliable measure, as we have heard over and over again. I think this is a global problem. What Paul

Watkins is proposing, a liver safety consortium, is another way to put some pressure on people to think about what they are doing, so we can interpret the results.

You have an enormous database in China. We also have an enormous database in the United States. The FDA gets clinical trial data from all the drugs, but of course, the data are confidential. So, they can't be released. But we have an enormous database of clinical trial data for real people that we can look at. That is what we hope to explore using the eDISH-2 program.

DR. GUO: I just wanted to add eDISH, actually, I haven't talked to anyone in China. eDISH actually is a tool. It can apply to all kinds of data. So, of course, we are dealing with clinical trial data that is confidential, but there is some room for collaboration, in my own personal opinion. So, we can apply it to some other data. That is possible.

DR. HONG: Yes, but my major concern with the eDISH system, it is still a very professional tool.

So, at this time only patients -- maybe trained doctors, we can use this system to collaborate. But in China a trained physician is very limited. So, we have been thinking maybe one day patients themselves can use this system rather than the physician using the system for diagnosis.

DR. GUO: Yes, this is something maybe we can talk offline.

DR. CHALASANI: That is a good idea.

DR. SENIOR: Maybe the patients could use for their measure, just a pin prick for the finger, and get at least a warning that there may be trouble, and get it right away. It doesn't have to be sent to a lab and wait until tomorrow to get the results. Get it right away.

Doing this for poor people in Thailand and Malaysia, and so forth, is something that is cheap. It is a postage stamp device that could be made cheaply, and it can give you results right away. Maybe it is not all that accurate, but it is close enough for clinical warning for patients alone, particularly in consultation with their physician.

So, educating the people, educating the physicians first, and then, educating the patients also will be, I think, important.

DR. CHALASANI: Thank you, John. I do think we need to move on, though, and then, you can take the offline discussions.

DR. QAZI: Nazia Qazi, here.Dr. Chalasani, you said that the guideline stated that you should consider stopping the medicine if there is evidence of liver dysfunction, which I completely agree with. But you also state that, if there is rapidly-rising ALT, you should stop it as well. Now in this very conference Dr. Seeff has mentioned adaptation. When does one think it's adaptation and when does one get scared about liver failure?

DR. CHALASANI: Once again, I thought you would ask me how do I define "rapidly rising". (Laughter.) I am glad you didn't. Go ahead. Sorry.

DR. QAZI: No, that was part of the question as well.

DR. CHALASANI: In clinical practice, though, I know if it is going from five times to

eight times, every time I see where something trends upwards as opposed to plateauing or improving. But in clinical trials, although the stopping rules now recommend expanding to five times and eight times, I know industry has been more receptive in trusting; the clinicians are the ones that are not. The minute it hits three times, there is a stop in the compound. It is not. Industry is really being very supportive in their stopping rules, in the protocols, et cetera. So, that is where I think people would disagree, you know, having a biochemical definition for DILI, but we are all over the place.

The international criteria say two times. I think the Barcelona says three times the upper limit of normal. That is Heithoff's paper as well. And the DILIN says five times the upper limit of normal. So, I think we are losing the window. If we can extend the stopping to five times, we may be able to stop who are adopting or adapting as opposed to who are progressing. Until then, I think it is going to be quite anecdotal, unless you

start seeing bilirubin or prothrombin time or symptoms.

DR. QAZI: The ALT is always muddy as well because people are on multiple medications and a lot of them have hepatotoxin potential.

DR. CHALASANI: That is really true. Yes?

DR. SEEFF: Maybe I will just make a comment about adaptation, which is a very difficult problem for me because I don't know how to tell when someone is going to adapt where someone is maybe on the way up. It seems to me that there is a three-part approach to this issue in clinical trials. Whatever level you set as the signal to have some kind of a concern -- and I know that companies would prefer to make it as high as possible, so as not to stop patients, and that there is a difference between the clinician and companies as to when they begin to get concerned.

But, to me, there are three parts to this. One is, if you set the level at three times the upper limit of normal, that does not say that this is drug-induced liver injury. It says something is

happening to the liver, and you don't know quite what it is. The idea, then, is to follow up and to repeat the test. I think, as part of that, you do the bilirubin as well. I don't think you do routine bilirubin unless people believe that that is the case. We routinely do ALT. But, if you have an abnormal ALT, then you do your bilirubin and, then, you begin to really get concerned. The question is, at what point can you tell whether this is an adaptation phenomenon? That is a common event. I mean, it is very common to have adaptation. We have seen this over and over again.

The only way you can do this is by following through. And then, it becomes an issue as to when do you stop the drug. If it goes up from three times to five times, to eight times? At what point do you decide to stop? So, perhaps there should be two levels. One is the signal to be considered about could this be drug-induced liver injury. The second is, is it real injury or is it simply enzyme elevation, because I think there is a distinct difference between.

And then, the question is, at what point in time do you consider the possibility of interrupting it for a time while you begin to look for the other causes? And then, if you find another cause for it, you can put the patient back on the drug and maybe it will come down. So, it is a very complicated issue, as I see it, at any rate.

DR. QAZI: I have one more comment. The recommendations on the statement of alkaline phosphatase elevated for a prolonged period of time and stopping drugs. As Dr. Dufour said as well, we have had patients in the clinic where the alk phos, when you fractionate it, you realize that it is coming from the bone. So, I think that should be kept in mind as well.

DR. SEEFF: Right. I agree.

DR. DUFOUR: The next question from this side of the room.

DR. MELLON: Yes, Eric Mellon from Pfizer. I direct this at Drs. Seeff and Senior. I certainly agree with your comments about the

narratives and the lack of completeness. I think, Dr. Seeff, a couple of years ago, you even made the comment, the biggest challenge for you in exonerating a drug is that there is often key information missing. So, I know from previous discussions many of the sponsors have hepatic data capture rates, but they all differ somewhat. And so, I am wondering whether it would be advantageous to try to standardize it. It wouldn't replace the narrative, but it certainly could help guide a narrative. You know, trying to standardize this sort of core set of information, particularly the things you find most often missing.

We all know the investigators, even GI investigators, I have some IBD programs I lead. When they get a case of potential liver injury, they are often a cardiologist or someone else who doesn't think about DILI all the time. So, I just wanted to know, would it help you if either you gave us some of the pieces of information that are often missing and critical to your assessment, and if industry tried to at least standardize a form. Our goal is

to have it as part of our standard case report form booklet, so that if they do get a question, in addition to the usual adverse event form, we ask them to go to that form and at least fill out the core information.

DR. SEEFF: I suspect that you are addressing this question to me in the belief that I still work for the FDA.

DR. MELLON: I did.

DR. SEEFF: And it is not the case anymore. I take care of a dog. (Laughter.) And it is a beautiful dog, and I am happy to show you pictures of the dog. (Laughter.) So, I would ask John whether he thinks -- when I was there, the big problem I had, as you say, was that the data that would come in were often totally useless. They were not being sent to us by a physician. They were being sent to us by somebody whose job it was to try and extract some information and send it out, and often it was totally meaningless. We had to go back and say, "We can't talk about whether this is drug-induced liver injury until you give us the following piece of

information." So, I think it would be a good idea to have some kind of an outline of what would be required in order for people at the FDA who try to assess these problems to be able to make some kind of a decision. And perhaps, John, if you wanted to make a comment on that, whether you think that an outline should be worthwhile?

DR. SENIOR: Well, I still do work for the FDA. They are kind enough to employ me at my advanced age. But, nevertheless, we are working on a revision of the guidance. Now the guidance was written in, published in 2009, the current guidance. It was written by a team of people, and Bob Temple wrote a lot of the language himself. We worked on that for two or three years. It is not easy to write guidance because you have to think of all the possibilities without making it too long and too complicated. How can you keep it simple and yet comprehensive? That is a difficult task, but we are working on it.

We have a team of people at the FDA: Mark Avigan and I and Lara Dimick and some others are

working on a revision of the 2009 guidance, which I think will address instructions not only to reviewers, medical reviewers looking at your new drug application, but also to clinicians who are looking at patients. That is a whole different world. Looking at clinical trials is one thing. It is an artificial world which is highly regulated. But looking at practice, patients in practice, not regulated. It is a whole different world. So, taking all this into consideration is our task.

DR. CHALASANI: That is great, John. The DILIN has published the minimal data elements required for a DILI diagnosis. I think Don Waki was the first author. That is evidence-based. I urge not to reinvent the wheel, but at least if you want to modify it, that is fine. But I really ask you to look at those. I think your point is well-taken. Everybody is moving toward this common definition elements, TDEs. So, I think maybe this evening you could discuss about having minimal elements that need to be in the reports.

DR. DUFOUR: We have about 10 minutes left. So, if we could try for short questions and short answers, we can hopefully get everybody.

DR. MA: Okay. Dr. Ma from China. I have one simple question about the biomarker cK18. My question is: what is the specification of the cK18, the diagnosis of DILI? Because the cK18 is also considered a biomarker for NASH.

DR. KIRBY: Yes, you know, I think the full-length is biomarker of necrosis. And so, I think it depends maybe on the threshold of concern. I don't know what the elevations you see with NASH. So, I think that is something that will have to be worked out with more clinical samples and experience. Maybe a DILI threshold for concern is different than a NASH threshold for concern. So, it is a good question.

DR. CHALASANI: I think it is very non-specific. I'm sorry. Because we just published a paper looking at K18 in PSC patients. It is high. Okay? So, I think it is all over the place. I don't think a threshold is going to make

much difference at all. This is, then, just talking about the K18 fragments, and we use the same kit as what you described.

DR. DUFOUR: The next question, on the right side of the room.

DR. TILLMANN: A related question to that, actually. So, my thought would be that perhaps there will never be a biomarker for DILI because DILI is so much different things. And therefore, perhaps we should focus our effort at the moment to find a biomarker to distinguish the people who have elevated liver enzymes and go on to develop a problem from people who do not go on to develop problems. And perhaps by not focusing on that same most important clinical question, we get the efforts diluted into finding a general DILI marker which perhaps never will exist because of the progressivity of DILI.

DR. CHALASANI: Hans, the way that will be answered is everybody continuing the drug in the face of increased ALT. So, that is the only way you would know who is going to go this way as opposed

to stay. Clinicians stop when you get three times, the majority of them at least.

DR. TILLMANN: No, but you would perhaps find markers which could lead to -- even though you stopped, you would have people who go on to develop severe injury, and you would develop markers to identify who would go on to injury. I recall a talk I think from the Liverpool group or someone from London where they had a case of acetaminophen, paracetamol-induced liver failure. I think it was HMBG -- which was in retrospect elevated at the first visit to the clinic when everything else looked normal.

DR. JESSNER: Wolfgang Jesnera from Janssen. We know that in viral hepatitis, and especially in hepatitis B, there can be ALT elevations, ALT flares, which are actually associated with efficacy, with e antigen seroconversion or s antigen seroconversion. So, we now are hoping in the next couple of years to develop powerful drugs tackling s antigen and cleaving to early s antigen seroconversion. It is expected that we might see

ALT flares in these studies as well. And I see a major problem here to differentiate between drug-induced liver injury and actually targeted toxicity. So, I wonder what your thoughts are to tackle that and, in particular, whether Dr. Kirby's biomarkers, if I might say, could have value in making this distinction.

DR. KIRBY: Well, in terms of differentiating those who may progress to DILI, I think some of the biomarkers that are sensitive to liver injury, they are going to go up if you have an ALT flare. But, if you are trying to see if they progress to DILI, I think some of the SAFE-T trials have that. They have enrolled patients from the DILIN Network, that maybe we can use these biomarkers to predict which patients will progress to DILI or not. Does that answer your question?

DR. CHALASANI: I think are we asking such a patient -- I mean, your drug, you are trying to get rid of surface antigen. If there is a flare, how would we know whether it is the drug as opposed to the flare, right? I think you just have to walk

through those cases and follow them. If those elevations resolve with the clearings of surface antigen -- I mean, otherwise, I don't know that it is now at a point at this stage we could. There are others in the room that are pretty experienced in this.

DR. JESSNER: Yes, but the question is whether these calls can be made at an earlier time, because of this other liver, let's say, petrogenesis, especially altering of hepatitis, as antigen withdrawal as well, but it is not so specific that we would like to continue the process.

DR. DUFOUR: Ask the next question on this side. We have about five minutes left.

DR. FRESTON: Yes, Jim Freston. Back to the practical question, how to write a proper narrative, we recommend people go on LiverTox, where Dr. Hoofnagle has actually provided information about what constitutes a proper narrative. He even gives guidelines of the elements that need to be included and shows

sterling examples. That might work for us until we get something from the FDA. But they are very good.

DR. CHALASANI: That's good. Thank you.

DR. DUFOUR: On the my left.

DR. HANSEN: Martin Hansen. I have a comment and, then, a question. A comment on Arie's comment on desirable context of use for new biomarkers, a point that came up repeatedly during the discussion. Actually, one of the originally contexts of use at the RMICT consortium was targeting was predicting who shows adaptation and who would is a susceptible patient. We had prospectives from a plan in patients on anti-leukosis treatment. We have included a lot of patients. I think, by now, it is over 200-and-something patients. But, actually, the incidence of DILI in that population was much lower than expected. So, we simply have to draw that context of use. That is another argument, I think, for going for larger stakeholder collaborations

and pool data, much larger patient populations. That is a comment.

And then, a question to Ted Guo and John Senior considering the further development of eDISH. There has been a publication, I think two years ago by GSK, using multiples of baseline and proposing a modified eDISH approach using multiples of baseline. They used a large non-liver-disease population of 16,000 patients. They came up with a recommendation on thresholds identifying potentially meaningful outliers. Does the FDA consider using multiples of baseline also in the future for the eDISH assessment?

DR. SENIOR: I don't think the answer is going to be multiples of baseline until we find out what the baseline is. We don't know that. We heard that from several speakers this morning. So, multiples of "X" are not useful. Now jiggling with the threshold, the cutoff on this is not the point. The chemistries are not the diagnosis. No matter how you jiggle the chemistries, it doesn't tell you the answer. The answer is medical

differential diagnosis of what is causing it. What is causing the liver abnormalities? It can't be done simply by adjusting the threshold, the cutoff point. I read the paper very carefully. It is not the answer. The answer is getting the information from the clinician on what was the cause of the problem. That is the question.

DR. CHALASANI: But, then, John, how do you compare the groups? I think maybe he disagrees with what you have said. I mean, everybody says precise diagnosis. But what you are asking is comparison between groups. That is really where the eDISH is, right?

DR. SENIOR: You have been working with DILIN for 10 years or more. You know that your primary goal in the DILIN, the network of experts across the country has been to make the diagnosis, right?

DR. CHALASANI: Part of it, yes, sure.

DR. SENIOR: Now you are not using eDISH. You are not just using RUCAM. You are using the collective wisdom and clinical experience of your

expert group. How can we capture that? How can we capture that wisdom to make the right diagnosis?

DR. CHALASANI: Are we using eDISH to diagnose DILI at bedside? No. We are using eDISH to look at studies. There is a signal in a program.

DR. SENIOR: It is being used for clinical trials, not for bedside, no.

DR. HANSEN: Maybe just to add on that, one of the crucial issues that we have is identifying liver signal, for example, in patients, and using the standard eDISH thresholds doesn't help us much in patients who have liver mets, or whatever. For those patients using multiples of baseline or any other --

DR. CHALASANI: Multiples of baseline, I don't think you get into a problem if it is within the normal, whatever, the low levels. My own view is multiples of the patient's baseline could work if you are starting with the higher levels.

DR. DUFOUR: Last question.

PARTICIPANT: Looking at eDISH in a different light and moving from the research realm

to the public health applications, why not explore using eDISH algorithms, coupling them with electronic medical records and commercial labs like Quest and LabCorp, and create an early warning system in the community for acute liver injury? Of course, you would have to screen out the population. But you could do that technologically now, and that might be one way to identify DILI earlier.

DR. CHALASANI: You know, a nice thought. I don't disagree. But the point is this: we have done an INPC in Indiana. Mayo has done it. You get a lot of false-positives, the reason being DILI is so rare. You are going to have a lot of cases meeting into this box or that box. You just have work though. But, absolutely, you know --

PARTICIPANT: That would also be an early detection for -- I know this is a DILI conference -- but it would help with early diagnosis of viral hepatitis, other forms of cirrhosis, PDC. Who knows? Just a thought.

DR. CHALASANI: Okay. I think that concludes this meeting session. It has been very enlightening and lively. So, thanks to all the speakers. (Applause.) 12:01 p.m.

Lunch break

Session IIA **12:58 p.m.**

DR. SENIOR: We had hoped to have Dr. Robert Califf here today, but when I invited him, he was still at Duke and he hadn't come to the FDA yet. So, he just arrived at the FDA on the 3rd of March, and he has been overtaken by an enormous amount of bureaucratic stuff. So, he is unable to be with us today. I am sorry about that.

However, we have a very good program. I want to start it off by introducing Lana Pauls, who everybody here knows. Lana has been absolutely wonderful in all of the things that she does to run these conferences for all these years. How many years have you been doing it, Lana?

MS. PAULS: Fifteen.

DR. SENIOR: Fifteen? How about that? She never misses a thing. We have about 180 people here from all over the world, from China, from Europe, from all over the world. We have practitioners. We have a lot of people from the pharmaceutical industry. We have regulators, and consultants. We have a wonderful assortment of people who we hope will be very influential.

What Bob Dufour said this morning is right on target. If we are going to get any improvement in standardizing the measurements of the tests we count on to make our diagnoses and our clinical decisions, we are going to have to do it ourselves. We can't wait for the federal government to pass new laws. We can't expect the bureaucracy to send out administrative fiats. But I think the practicing doctors and the consultants who are the the hepatology experts. There are many of them in this room. They consult both to industry and to the FDA. And they are very influential in steering things around. So, we have to make a case for the importance of what we say here.

We will try to capture everything that is said in the discussion, and we appreciate the excellent discussions we have had and will have. That discussion is captured word for word. A transcript is made; it takes a couple of weeks for the court reporter to send us a draft transcript. He goes over it very carefully. And then, we edit it even further to make sure it all makes sense before it gets put on the internet.

Once it is on the internet, it is open to the whole world. It isn't just this year's conference. What is on the internet goes way back, way back, all the way to the first conference which was in April 1999. That was 16 years ago, the first conference. It was a conference only for FDA reviewers, but we had 400. Bob Temple and I were there. Bob Temple at that course was one where I think he first gave his definition of Hy's Law, right, Bob? Yes, something like that. (Laughter.)

Anyway, without further ado, I want to introduce someone that you don't need to be introduced to, Lana Pauls. (Applause.)

Pauls photo, biosketch and abstract

LP#1: Thank you, John. I have the distinction of starting off this session. Really, I am only going to be talking for about five or six minutes, just to set the stage so you can be aware and think of some of the other speakers in this session. I am one of the only non-clinicians here that will be speaking today, but I am doing this from an aspect of being here for the last 15 years and listening to a group of you very intently over the last 15 years and learning a lot.

So, with that, I was charged with looking at this from a very different perspective, literally talking about whether it is the drug, the chemical, the hepatotoxic agent, and/or the person. As I said, the theme of this session is really all about understanding one another. I am going to present some basic concepts. And then, John is going to focus on what we have to do to get it right. And then, Dr. Alice Chen is going to focus on some more

clear definitions and terms which we have been struggling with for numerous years.

About 10 years ago, I was involved in one of the subcommittees on nomenclature, and we still never finished that. So, that is another issue that we have on a regular basis.

Drs. Carter and Hicks will focus on the true causes and various disorders associated with this. And then, Dr. Temple will close out the session talking about some labeling concerns that we have.

LP#2: Here's my standard FDA disclaimer. I want to start out with some known facts about DILI. And I am speaking to people that know this very, very well.

LP#3: We all know that DILI is a major concern in medical practice and in public health. We also know it is one of the leading causes of acute liver failure in the world. We do also know it is a major cause for drug failure in clinical trials.

If you look a little bit further down here, up until 1997, it was actually the leading cause for withdrawal of any drug from the market. That

year of 1997, FDA approved eight drugs later withdrawn, four of which were for liver failure.

We also know it is a very rare problem. So, it is very hard to find in clinical trials. Anytime we see an incidence of it or a detection of it, we really are concerned about it.

And lastly, we all know that, associated with the adverse events, they are severely, severely underreported. So, we really don't know the actual incidence. Liver failure is associated with not just prescription drugs, but it is also associated with other agents, including over-the-counter medicines.

LP#4: Dr. Lee was not able to be here this year. So, I felt compelled to show his data because this slide gets shown at every single conference. This slide shows data through 2014. As you can see, 974 cases were associated with acetaminophen.

LP#5: Here is the list of drugs that, again, were approved in 1997, four of which came off of the market that year.

LP#6: So, moving to the fundamental questions associated with DILI, is it the drug that is toxic or is it specifically a susceptible person? Well, actually, it is usually a little bit of both, and sometimes it is very, very difficult to discern the difference there.

LP#7: So, even though there may be a safe dose for most people, it is not necessarily safe for all. I know for a fact that, when I speak on behalf of the FDA, quite frequently, people that are unfamiliar with the way that we do business often think that just because a drug is approved that it is has kind of, sort of, got a Good Housekeeping Seal of Approval and it is safe for everybody. But, as clinicians and other healthcare providers, we know that that is not necessarily the case.

A couple of the other facts associated with DILI: As I indicated, the same drug might be quite safe for most people, but it is toxic for a number of them. And it is associated with different severity, consequences, what the time course is.

And when I say "serious," this can lead to a minor disability, an inability to work, hospitalization, liver failure, or in the worst case even death.

We also know that the liver has an amazing capacity to recover from injury. The fact that you can have it two-thirds resected and it can regrow, even if some of the hepatocytes are killed or removed, and it is very, very adaptable.

LP#8: So, one of the problems associated with this is the identification of what is going on are very challenging. We have to look at the dose and the properties of the drugs that impact the initial cellular damage. We have to look at the host factors that drive susceptibility of the agent as well as the repair of the liver, and we also have to look at the fact that primarily this is idiosyncratic.

LP#9: Of course, this conference wouldn't be this conference if we didn't mention Dr. Zimmerman on a number of occasions. John, of course, just mentioned Hy's Law that goes back at least 25 years. Dr. Zimmerman had a lot of different ideas, one of

which he talked about was that liver injury is also associated with, again, not just drugs, but other substances, including plants and animals. So, you can think of things like, other things that cause these injuries. And he always talked about these along a spectrum of toxicity.

LP#10: So, there are also risks in humans that are likely to be determined by multiple factors, including the drug properties, the patient attributes, and the various DILI mechanisms. Nobody ever said that this was easy, and that is probably why we convened this workshop every year for the last 15 years.

LP#11: Some of the different drug properties related to DILI include threshold dose, lipophilicity, reactive metabolites, oxidative stress, and mitochondrial liability. I am not going to go into those any further because this is going to be the primary focus of Dr. Chen's talk tomorrow afternoon. So, be looking for that.

LP#12: We also know that predicting serious liver injury has its challenges. We know that

biomarkers are not necessarily specific enough. We also know that negative rechallenge can be unconvincing, especially for rare events. We know that positive rechallenges is very powerful, but in some cases it can be very dangerous as well. And lastly, we know that it is very difficult to determine the causality of this.

LP#13: Again, over the last 15 years, I have been listening to all of you and I have been learning a lot. I have always wondered myself what makes the certain people respond better to the same dose and regimen of the same drug than others, and what makes certain people susceptible to the serious adverse events associated with this, when most people aren't?

LP#14: So, all I am asking now, at the end of this talk, is that you listen carefully to the speakers who follow me. They are going to be addressing a lot of the topics that I highlighted in this opening session. Thank you. (Applause.)

Senior photo, links to biosketch and abstract

JS#1: DR. SENIOR: Thanks, Lana. Well, we have been talking about whether we understand each other? Leonard Seeff this morning said he has been saying the same thing over years and years. Bob Temple has been saying the same thing. He points to the 2009 guidance. He has said it very clearly; it is all there. We keep saying the same thing, but people don't seem to hear it.

Thinking that maybe speech isn't the best way to communicate, maybe writing is more important. Write it down, and maybe we can get it better understood. So, I thought I would do a little experiment.

JS#2: Lana said the theme of the conference is really about understanding each other. When we write something, we take trouble to try to write it carefully. Also, if we are going to get it published, it goes out and it is reviewed by our colleagues, who give it a tough review and they will criticize it. They won't accept it if it doesn't make sense. Then, it is also subject to editorial commentary and review. So, by the time it gets

printed, it has been through a lot of critique. So, we try to write it carefully. It gets looked at carefully before it gets published.

JS#3: Okay. Once it gets published, it is in the literature. Well, that's nice. About two years ago I was asked to write a chapter for a new book on antitargets. Now what is an antitarget? An antitarget is a site, a receptor, that is affected by a drug that is not intended to be affected. So, it is what is causing the adverse effects. The target is the therapeutic effect. The antitarget is the unwanted adverse effect, which is very unpredictable. So, this is a book on antitargets from all aspects. I was asked to write about the clinical aspects of liver injury from antitarget effect, from effect on receptors not intended when the drug was being developed and is being used for its therapeutic value.

JS#4: So, I figured that we want to look at the past so we don't repeat the same old mistakes over and over again.

JS#5: So, what I did was a little experiment. I wrote the paper and I went through a number of revisions. I cited 52 references, but I took the trouble of getting a copy not only of the abstract, not only the title, but each full paper. I got the full paper PDF using our nifty FDA library. I got all 52 citations word for word and went over them one more time. And then, I sent the paper to the person whom I cited and along with what I had said about them and said, "Did what I say correspond to what you wrote, what you intended me to understand? Did I understand you and did I get it right?"

So, I sent them out. I couldn't find email addresses for everybody. Many of the people I cited are not with us anymore. I was only able to get 19 email addresses of the 52 from authors who have published more recently.

JS#6: And so, this is what I said to them in an email back in January: "We would like to do this little experiment. I send you a copy of what you wrote some years ago; I want you to look at it one more time, and see if what I said correctly

corresponded to what you intended me to get from your message." So, it was aimed at understanding each other about what we write.

JS7: I said, if we don't get any responses, we will call it a zero. If we get a response from the author of the paper, "Right on, you got it right; you understand," that is a one, and so on down. Two, no objection. Three, maybe I didn't quite understand it. Four, you're absolutely dead wrong. Five, maybe we should have explained it more clearly, and other comments.

JS#8: Here are the results of the January mailing. More than half didn't respond at all. Oh, dear. Only four people out of 19, or about 20 percent, said, "You understood what I said." I didn't really learn much.

JS#9: So, I said, this was a waste of time, but I will try again. I dug out 10 more email addresses, to bring the total to 29 of the 52. So, after that poor response, I decided I would try again in February, a month ago. I didn't send it

to people who had already responded, but to about 20-some more different people.

JS#10: And this time, out of the 29, we got 10 people who said, "Yes, you understand what we are saying. You got it right." But we still didn't get responses from a lot of people.

JS#11: So, the conclusion was this was a bust. (Laughter.)

JS#12: This was not a good idea. And I said, well, I haven't failed, but I have just found a lot of things that don't work, as Thomas Edison had said. So, even the brilliant minds of our scientific world have made mistakes.

JS#13: We can are making new mistakes; not just repeat old mistakes. There are lots of opportunties to make mistakes in this world. So, I don't think I will do this again. (Laughter.)

It was an experiment, but it was a lot of work. A lot of people, frankly, are too damned busy, but. I only heared from one person who said, "I won't respond to you because I'm too busy." (Laughter.)

The rest were good people, just didn't respond at all. Nevertheless, I have a message: It is a good idea to read papers carefully; It is a good idea to be careful what you write; It is even a better idea to do that before you send it in for publication.

There's lots of room for mistakes. We all make mistakes, and we shouldn't expect always to get everything right the first time.

JS#15: And so, I close with an example of what Lana has just shown you, an example of a group that is called the Acute Liver Failure Study Group, headed by Will Lee in Dallas. Looking year by year, now for about 15 years he has been looking patiently at the incidence of acute liver failure, and he keeps getting the same results.

Moderator Session IIA - 2

There is an enigma here: How can acute liver failure be caused by drugs more than by anything else, when we all know that acute liver failure from any given drug is rather rare? One in 1,000, 1 in

10,000. If it is the most common cause, and yet it is so rare, how is that explainable?

Well, probably the answer is, for any one drug DILI is rare, but people take so many drugs and are exposed to so many chemicals that the combined effect is not rare. It is not just prescription drugs; it is from all the over-the-counter medications that people take, and includes the generic drugs for which you don't even have to get a prescription. You can just go to a grocery store and get them, and it includes herbal and dietary supplements which are not regulated at all. It even includes chemicals such as alcohol. Alcohol is a drug. Did you all know that? But it is not counted as a drug; it is not classified as a drug; it is not classified as a dietary supplement or a nutritional aid, but as a “taxable commodity.” The federal government classifies it that way just to make revenue, but refuses to take up the challenge of calling it what it is. It is a drug that causes probably more liver injury than all the rest of them put together, but only in some people, not in

everybody, not even in people who are heavy drinkers, not even in people who have a drink every day, like I do. (Laughter.)

Particularly, as Daniele Prati said, it is the binge drinkers that get in trouble, not all binge drinkers, but a lot of binge drinkers. Some years ago I was at the Graduate Hospital in Philadelphia, where I went to run the Clinical Research Center. It had its money withdrawn by NIH, so my funding source was gone. Well, the hospital offered me a job, saying: "We have a proposal from the Diagnostic and Rehabilitation Center of Philadelphia." The DRC included Philadelphia and six counties surrounding, four in Pennsylvania and two in New Jersey. DRC said, "We need a hospital to send alcoholic patients who are so sick that they can't be cared for adequately in their local hospitals." So, if the local hospital felt that the patients were so sick that they might die, then they would send them to Graduate Hospital, to me. Well, over 4 years I admitted 3500 alcoholics with severe, life-threatening medical complications.

I did not try to treat their addiction; I was just trying to treat their bodies, to get them over their heart failure, pancreatitis, liver failure, their neuropathies.

We had counselors coming in from DRC who worked with them when they felt a little better and could lift up their heads and listen. So, the counselors worked on the addiction; I just worked on the medical aspects. And I learned a great deal about the problem of alcohol-induced liver injury. And, yes, it was people who tended to be binge drinkers who got into trouble. After an attack of acute alcoholic hepatitis, they would then say: "I won't drink anymore." That lasted for maybe a couple of weeks, and the, they would relapse. Have you seen that before? Yes, you have; we have all seen it. They relapsed; they couldn't stay off the sauce; They were addicted. I had one patient who I treated 13 times. He was a PhD and a most well educated, most literate person, but he couldn't stop drinking. It took 13 visits for him to come into our unit in liver failure before he finally

got the word to quit drinking permanently. And he has recovered.

But you can't measure alcohol injury simply by saying, "Well, anybody who takes more than two drinks a day is at risk." It is not so. A lot of the heavy drinkers never developed liver disease at all, even the binge drinkers. Have you seen that? Lots of people can tolerate a lot of alcohol. They don't get liver injury. It is not automatic, or just dose-related, but has to do with the individual susceptibility. These were lessons learned from this experience.

And what Will Lee has shown I think is a lesson we have all learned. Do we really understand what his slide means? Do we really understand that acetaminophen plus prescription drugs, 46 and 11 percent, that is 57 percent, cause more acute liver failure than all of the liver diseases put together? Isn't that remarkable? And that is not just true in the United States. This is being confirmed all over the world. Drugs and chemicals are causing more liver failure than anything in the

whole world, more than all the diseases. So, that is why what we are doing here at these meetings Is very important.

I compliment Will Lee for sticking with it and for publishing this graph every year. There were 2100 patients last year, now over 2200 this year; it will be 2300 or 2400 next year. It is consistently telling us that acute liver failure is caused by drugs and chemicals, and it is our job to find out why and how, how to detect it, how to deal with it, how to treat it, when to take the drug away, when to continue the drug. This is what we have to do. This is the challenge to all of us.

So, I thank you for the comments, and I will turn this over now to Dr. Alice Chen. Let me say a word about Alice Chen. There was a Fogarty Conference down at NIH in 1974 that I attended on nomenclature of liver disease. That was followed in 1978 by a second Fogarty Conference which was on setting standards. I didn't get to that because it was when I was taking care of all those people in Philadelphia at the Graduate Hospital. I think

Bob Temple was at that Fogarty meeting. Is that right, Bob? And he heard the discussion from all these leading experts from all over the world. And he was very impressed with what Hy Zimmerman said. Hy would say over and over again, "If you see drug-induced hepatocellular jaundice" -- three things. It had to be drug-induced, not caused by disease; it had to be hepatocellular, not biliary; and it had to be severe enough to cause jaundice, not just some transaminase elevation, but dysfunction of the whole liver. But drug-induced, hepatocellular jaundice has like a mortality of 10 percent or more, which has been confirmed over and over again. That is what Bob heard. It was not written in the meeting summary, in the book that summarized the Fogarty Conference. But it set standards. At that meeting it was the consensus, not from data but by opinion, that threefold elevations of transaminases are markedly abnormal, whatever that means. That is all they said. Threefold elevation is markedly abnormal. They

didn't say how many units. They didn't say anything else, and they didn't say what it meant.

But the idea was picked up by NIH in the first series of clinical toxicity criteria, 1982 I believe or 1983. And they have been continued since and updated. Dr. Alice Chen at NIH is now chairing the revision for Version 5. We are going to hear from Alice Chen about criteria to be measured for laboratory tests and the new common terminology criteria for adverse events. Dr. Chen?

Alice Chen photo, links to biosketch and abstract

AC#1: Thank you. I was a little worried that Dr. Senior was going to say that I met him in 1974. I'm trying to remember how old I was then. (Laughter.) But I want to thank him for inviting me to participate in this meeting.

I am a medical oncologist. And so, my whole world kind of revolves around, when you talk about drugs, it is always oncology drugs. I just want you to have some perspective of where I am coming from.

How many people here actually know what CTCAE is? (Show of hands.) So, I will probably go through a little more in detail. CTCAE is a document that primarily is used to assess adverse events in terms of drug development.

The Cancer Therapy Evaluation Program of the NCI in 1983 developed this in order to kind of cross the multiple trials that they were supporting to have more uniformity in terms of adverse events that were coming in.

Because if we just left it up to everybody, somebody who is bleeding, could come in as low platelet counts, thrombocytopenia, molar suppression, different terms. And so, in order to have some uniformity, we developed a guide in terms of the common nomenclature and how these adverse events need to be reported. That has been revised three times now. Version 4 was released in 2009. With that there was a huge change in the CTCAE. Up until then, our adverse event terms were done independent of any documents. From what we have seen in our reporting by our investigators, we have

picked up those terms and have listed them under categories.

AC#2: But we then realized that a lot of the pharmaceutical companies were doing their adverse event reporting to both the EMA and the FDA using MedDRA. So, we got involved with MedDRA around 2008 and started to change all the CTCAE terms to MedDRA terms. We are hoping that that allows it to be a document that is easier for people to use, especially as everything becomes more electronic. The categories become System Organ Classes (SOCs), so that where these AE terms come in is determined by MedDRA to some extent. And I have to admit that we have gotten some of the comments since the release of 4 wanting to know where to find these terms from people who were not accustomed to MedDRA and didn't understand why certain terms are put under various SOCs.

Also, when we released in May of 2009 CTEP, then, systematically, they support thousands of trials. They systematically start to convert all our trial reporting to CTCAE Version 4. Now this

is not true for all the trials out there because some of the PhRMA-sponsored trials continue to use older versions, but we couldn't support Version 2, 3, and 4. So, we have just made a decision to actually convert all the trials, as of September of 2010, to use CTCAE Version 4.

There was also a core group that was developed at that time to allow us to continue to look at the comments that came in for CTCAE Version 4 to see if there were any mistakes that were made or in terms of when we need to come up with a new revision.

AC#3: So, these are the SOCs, the System Organ Classes of the MedDRA terms in which we have placed the CTCAE adverse events for Version 4. This is kind of what the document looks like.

AC#4: I just selected the hepatobiliary disorders. You have the adverse event, and then, there are five grades. That is one of the reasons both companies used the MedDRA term for reporting as well as CTCAE, because this does provide grading.

AC#5: How the grading occurs, Grade 1 is usually asymptomatic, something that is just found because you did a test. Usually no intervention is required.

Grade 2 is usually there is some intervention that is required, does not require hospitalization. Usually, Grade 2 in most of the trials for therapeutic, not so much in the prevention setting, for Grade 2 we do not change the doses.

Grade 3, the big difference between Grade 2 and Grade 3 is that for Grade 3 a lot of times it requires stopping the drug or clinical trial. And so, when we did 4, we try to keep that in mind as we come up with the criteria for Grade 2 and Grade 3 toxicity. Grade 3 usually requires hospitalization, IV, or some type of surgical intervention. Grade 4 is immediately life-threatening, and then, Grade 5 is death.

So, this is just to go through the process in terms of how we are going to go about with coming out with Version 5. And so, we had monitored the comments. There was a help desk email since Version 4 was

released, and we have taken all the comments and went through them. We have a contractor to assist us in terms of managing all these comments that came in; personal communications from our investigators as well as investigators within CTEP, as they are looking at these adverse events coming in, if there are ones that really are difficult or the grading, there is any problems with them.

The core group has met as needed for these comments, and we have reviewed the impact. So, one of the things we did in April of last year was we released the draft version for public comments. I'm sorry. Actually, we have let everybody know that we were going to revise CTCAE Version 4 and asked for public comments.

And we got a lot of comments, which delayed the Version 5 release, because we had to process all those comments. How we went about these comments is that we looked at, especially if there are new AEs that they want to add, we wanted to know how frequently it was actually reported. Usually, those will be reported as "other".

So, under CTCAE, each of the categories or the SOC, there is an "other". So, if there is not an appropriate term for a certain adverse event, they can report it under "other," and we have pulled all those up to see if any of them occur frequently enough for us to add as a new event.

There has been some confusion in terms of grading, and we try to clarify that. One of the ones is that, for creatine, actually, Grade 1 is anything above baseline. As you know, if you measure your creatinine today and tomorrow, probably half of you will be above what you were yesterday. Unfortunately, nobody actually picked it up until this year, but we will be changing that.

Clinical significance. As to how we manage adverse events and new drugs come out to manage some of our adverse events, the grading sometimes has to be modified because we manage our adverse events better.

The last thing is that any new term has to be a MedDRA term. We have consulted with our Working Group members from Version 4. We actually had a

Working Group for each of the SOCs, including experts within that area, to help us in terms of managing and making sure the terms are appropriate in terms of the grading and management of the adverse events, NIH members and academic experts.

AC#6: So, what is not going to change for Version 5 is that the SOC and the term placement within the SOC stays the same. That is still driven by MedDRA. They are all going to be MedDRA terms. We are still going to stick with five gradings. The guideline for each of the grades does not change. There is no deletion of any AE term. So, one of the concerns is that, if we delete a term and a study takes longer than nine years to do, then that adverse event does not have anywhere to go when the study is reported. So, we are not planning to delete any of the AE terms that are currently in 4.

In the past when we have changed our AE terms between 2 and 3 and 3 and 4, we have produced a mapping document to lead people from one adverse event to another, or if we change any grades, but

we are not planning to provide a mapping document this time because they are all MedDRA terms. So, we are planning to delete any terms.

We have a lot of comments for the use of upper limit of normal or lower limit of normal for these lag values. We are planning to keep that because of, I think, some of the things that were discussed this morning.

There is no uniformity in terms of -- you know, for a white count or WBC or ANC, there are set values that everybody uses in terms of managing toxicity. For AST/ALT there is really no set values right now. It is set by the labs. And so, we are going to keep the ULN and LLN.

AC#7: What may change in CTCAE Version 5 or what will change is that we have new AE terms added at the recommendations of the public, clarification of certain definitions, add or clarify and change in grading, some editorial changes that were never picked up, despite multiple layers of review.

We are also going to add navigational notes. That was taken out in 4. In 3, those of you that used it, there were things that said, "also consideration" to help in terms of managing different similar AEs, and we are to report them. So, we are, for like ALT, which I will show a little later, we are planning to have a navigational note in terms of considering hepatic failure. We are also going to provide an index so people know where to find these terms.

Though currently CTCAE is online and it is an Excel spreadsheet, so that you could actually use a Find function to search any AEs.

AC#8: So, just examples of new AEs: disease progression was done; it is not an adverse event, but it is added so that it is helpful in terms of tracking. Some of the hepatitis B reactivation will be added as an adverse event because of some of the immunotherapy agents that are out there. Budd-Chiari syndrome will also be added as a new term.

AC#9: What will be changed in terms of definition? None of these really are in terms of the DILI, but if there is any in terms of the definitions that you felt needs to be changed, please let me know.

AC#10: Added or changed grades, the biggest one is actually the last one. So, neonatal death, because it was a death, it was listed as Grade 5. The problem is the patient is still alive, and the computer system does not allow patients to go forth in terms of further treatment. So, we are going to change neonatal death into a Grade 4, so that patients can continue on treatment, if needed.

AC#11: Grade clarifications. We added various things to clarify some of the adverse events.

AC#12: And then, navigational notes. If you will look at the yellow part, the AST, ALT, and bilirubin, we are going to add a navigational note to those because those are lab values and only use specific numbers.

AC#13: We are asking that you also consider hepatic failure, if appropriate, for reporting. So, these are the adverse events in CTCAE Version 4, and the yellow ones will be added to 5.

AC#14: I just wanted to talk about the investigational SOCs. So, a lot of the first sign in terms of any adverse event is laboratory values. Unfortunately, it doesn't always reflect what is going on in the patient if you use an absolute value. However, that is a way that is common and easy for us to assess for these.

AC#15: So, everything that is an investigational SOC is predominantly driven by numbers. Usually, we use upper limit of normal because of lack of a standard value, but we selected for the ALT is greater than three times upper limit of normal.

AC#16: After discussion today, actually, one of the considerations is if we should consider a change from baseline as well. If you note, there is no Grade 5 in that we don't think anybody can die from a laboratory value. They die from a

medical condition. So, we want them actually to report the Grade 5 under the actual condition in which the patient's death had occurred.

AC#17: So, just areas of consideration for our discussion for today, alkaline phosphatase, though there is a Grade 3 and Grade 4, because Grade 3 and Grade 4 is very important in terms of some of the oncology trials for stopping patients' drugs, we actually would like to change those to just Grade 2, because anything above that that leads to hospitalization or immediately life-threatening should really be reported under the actual medical condition. GGT is the same thing. We are proposing to convert the pure values, everything down to Grade 2. And then, for everything above that, to actually be reported under the medical condition.

And the question is if we should consider that for both AST/ALT, and I guess for bilirubin. I would certainly love to hear your thoughts in terms of that. Any other AEs that we need to include. With recent use of immunotherapy as well as some of the targeted agents, there have been new adverse

events, and we want to make sure we include those new AEs as they come up. And then, any other CTCAE changes that would assist in better assessment of DILI.

AC#18: The CTCAE, the link is to Version 4. And then, I guess the last actually also has the email address in terms of if there are any comments. Thank you. (Applause.)

Moderator Session IIA - 4

DR. SENIOR: Previously we were considering what does a drug do to a normal liver and if you can detect an injury caused by a drug? However, recently, we have become concerned there are certain conditions where the liver is not normal before you start treating. Yet, you still are worried about the possibility that there may be drug-induced injury on top of the disease.

We are going to discuss now two such prominent disorders. One is viral hepatitis. You are all aware of the work on hepatitis C, the multitude of new treatments, some very effective for treating

and even curing it. So far, we have not found that any of the drugs used to treat hepatitis C are causing DILI, but we have to be vigilant because new drugs and new combinations are coming all the time. The second condition for which abnormal liver function pre-exists is chronic heart failure. Now we all know that the kidney and the brain dysfunction are secondary to liver dysfunction. If you get liver failure, you get renal insufficiency and encephalopathy.

Perhaps we have perhaps forgotten that the liver is dependent on the circulation of blood to it by the heart. So, if you have heart failure, you get secondary liver failure. Now this is particularly bad in the situation called shock liver or ischemic hepatitis, as was so forcibly introduced by Maddrey and Boitnott some years ago. In patients who go into shock and recover, the serum liver enzymes may go sky-high, 30, 40, 50 thousand units per milliliter, but they come down very quickly if you restore the circulation of blood to the liver. The liver is very dependent on oxygen;

it uses a lot of oxygen to do its metabolic work. And if deprived of that oxygen, liver cells die and they release enzymes into the plasma.

In congestive failure, it is a different situation. It is not that the blood is not getting to the liver; it is that the blood can't get out of the liver and back to the heart because it is backed up. And so, the liver, being hungry for oxygen, uses it all up and the centrilobular regions, which are at the end of the oxygen supply, tend to show necrosis. So, both congestive failure and shock liver cause secondary injury to the liver, but it is really due to the heart disease. And you don't treat the liver; you treat the heart. So, we are going to hear first from Karen Hicks, who is going to be talking about congestive heart failure. Karen?

Hicks photo, links to biosketh and abstract

KH#1: Good afternoon. Thanks, John, very much for the invitation to be here today. I am delighted

to discuss elevated transaminases in the setting of heart failure.

KH#2: I have nothing to disclose, and the opinions expressed here are my own.

KH#3: The objectives of my presentation today are to review the effects of acute and chronic heart failure on acute and chronic liver injury, respectively, and also to discuss how challenging it can be to assess the potential for a drug product to cause drug-induced liver injury in the setting of heart failure.

KH#4: As John has mentioned, there is a mutual relationship between the liver and the heart. Hepatocardiac diseases can be divided into three categories: heart diseases affecting the liver such as heart failure, which will be the focus of my discussion today; liver diseases affecting the heart such as chronic hepatitis C, which will be discussed by Dr. Wendy Carter to follow, and conditions affecting the heart and the liver at the same time. If you would really like to find out more about all three of these

conditions, I recommend the review article by Fouad which was included in the references.

KH#5: As we all know, acute heart failure can lead to acute liver injury, also known as acute ischemic hepatitis, and chronic heart failure can lead to a chronic congestive hepatopathy, also known as a nutmeg liver. But, as John has mentioned, the goal is to treat the underlying heart disease, the underlying heart failure, because usually the hepatic abnormalities will improve.

KH#6: Well, what in the world is a nutmeg liver and how did we ever come up with that term? I am not a histopathologist. But, if you looked at the picture on the right, those are normal liver cells. You see a picture of the nutmeg on the left. If you remember only one slide from my presentation today, I hope it is this. And if you play Pictionary, you may want to keep this in your back pocket.

KH#7: But a nutmeg liver is not a happy liver. This is what it looks like. The term "nutmeg

liver" was coined by Kiernan in the 1830s and, subsequently, by two others in the 1870s, by looking at autopsy specimens. And then, Mallory came along in the 1900s and determined that these brownish-red spots throughout the liver are actually due to centrilobular necrosis. I should say that medical history could have been very different if all three of these individuals had thought about cinnamon first.

I don't know what it is, but after being here at the agency for over 11 years, I find that sponsors, in particular, don't really know when the FDA may be trying to make a joke. (Laughter.) I'm glad I got your attention.

KH#8: So, let's be very simplistic. Liver injury can be divided into acute liver injury, acute ischemic hepatitis, or chronic liver injury. Let's talk about acute liver injury first. As John has also mentioned, the liver has a great capacity to compensate for significant insults such as hypotension and decreased blood flow and hypoxemia.

KH#9: What it does is try to extract more oxygen from the blood that does flow through the liver. But sometimes even that compensatory mechanism can be so overwhelmed, and you end up with hypoxic damage and get hepatocellular injury.

KH#10: There are three main insults that can result in acute liver injury. The first is hypotension. The second is hypoxemia. And the third is increased metabolic demand.

KH#11: Profound hypotension can be caused by acute cardiopulmonary arrest such as in the setting of acute myocardial infarction. It can also be caused as a result of heart failure in and of itself, but also associated with acute myocardial infarction, like a Killip III or IV infarct. You can also see hypotension with pulmonary embolism or sustained arrhythmia such as atrial fibrillation or flutter with a rapid ventricular response. Observationally, heart failure accounts for most cases of acute liver injury.

KH#12: Hypoxemia can be due to respiratory failure or obstructive sleep apnea. With all of

the obesity that we have in this country right now, obstructive sleep apnea is a big problem. There is a lot of people on CPAP.

KH#13: Toxic or septic shock can also cause increased metabolic demand and contribute to acute liver injury.

KH#14: What are the signs, symptoms, and physical examination findings that we see in acute liver injury? The patient can either be asymptomatic or have some non-specific symptoms, such as nausea or vomiting, anorexia, malaise, right upper quadrant discomfort, jaundice, decreased urine output, or flapping tremors, which are due to cerebral hypoperfusion and not hepatic encephalopathy.

KH#15: With respect to the laboratory evaluation, we typically see sharp increases in the transaminases, total bilirubin, alk phos, LDH, and PT, occasionally accompanied by renal impairment due to acute tubular necrosis. These liver abnormalities typically peak one to three days after the onset of the insult and normalize within

five to ten days. You will also typically see an ALT/LDH ratio of less than 1.5. That helps to differentiate acute liver injury from viral hepatitis or even DILI.

KH#16: Pathophysiologically, what we see is centrilobular necrosis of the zone 3 hepatocytes. I am not a histopathologist, but there are three Rappaport zones in the liver.

KH#17: The central vein is located in the central part of the lobule, which is poorly oxygenated, as opposed to the portal tract where the hepatic artery runs.

KH#18: And so, zone 1 is highly-oxygenated. Zone 3 is not and is most susceptible to anoxic and hypoxic injury.

KH#19: Now let's move on to talk briefly about chronic liver injury in the time that we have left. Chronic heart failure and hepatic dysfunction can be due to a number of conditions, such as ischemic or non-ischemic cardiomyopathies, pulmonary arterial hypertension, valvular heart disease such as mitral stenosis or tricuspid regurgitation,

constrictive pericarditis, and postoperative consequences of the Fontan procedure.

KH#20: What are some of the symptoms and the signs and the physical examination findings we see in chronic liver injury? Well, you could have mild, dull right upper quadrant pain, hepatomegaly, peripheral edema, ascites, and jaundice, although jaundice is uncommon. Ascites can occur in up to 25 percent of patients.

KN#21: From a laboratory perspective, you will have two- to threefold increases in AST, ALT, LDH, GGT, and ALP. The total bilirubin will be increased, but it rarely exceeds 3 milligrams per deciliter, and the albumin can be low.

KH#22: Pathophysiologically, what we have here is we have right ventricular dysfunction which increases venous pressures. You end up with atrophy of hepatocytes, perisinusoidal edema, increased lymph formation, thrombosis due to the stasis of the blood flow within the sinusoids, the hepatic venules, and portal tracts. And you see an alternating pattern of hemorrhage and necrosis

in zone 3 and normal or slightly steatotic areas in zones 1 and 2.

KH#23: So, if we compare acute liver injury versus chronic liver injury, the point I just want to make is that, typically, in acute liver injury and acute heart failure there are marked increases in the total bilirubin and the transaminases compared to mild increases in the setting of chronic heart failure. With respect to acute liver injury, usually it is benign and a self-limited course. With chronic liver injury and chronic heart failure, if you treat the underlying heart failure, in the near-term the liver abnormalities can improve, but overall there is going to be a slowly progressive course due to the chronic heart failure.

KH#24: So, to contrast heart failure versus drug-induced liver injury, this has been covered by all of our other speakers this morning, including that very interesting talk by Dr. Seeff, I am sure as you all know. Where is Dr. Seeff? Did he leave? I did not find his lecture to be boring.

So, what you get with DILI is hepatocellular injury, elevations in the aminotransferases of greater than threefold upper limits of normal, total bilirubin greater than two times upper limit of normal, a normal alk phos, and there can be no other reasons to explain these abnormalities. In many cases, DILI is not dose-related or evident non-clinically. In some cases, it can be idiosyncratic.

KH#25: In summary, there is a mutual relationship between the heart and the liver. Acute heart failure can lead to acute liver injury and acute ischemic hepatitis. Chronic heart failure can lead to chronic liver injury. You want to treat the underlying heart failure. The liver abnormalities will typically improve. In the setting of heart failure, it can be very challenging to assess whether a drug product can cause drug-induced liver injury. That is why I want to say that I am so grateful to have a colleague like Dr. Senior who can help us sort through these

very challenging cases. Thank you very much for your attention. (Applause.)

Moderator Session IIA - 5

DR. SENIOR: Dr. Wendy Carter has been working with the Antiviral Division at the FDA for several years and has been through a number of very significant approvals of exciting new drugs that approach almost a cure for hepatitis C. Wendy?

Carter photo, links to biosketch and abstrract

WC#1: They are a cure, yes. The good news is that a lot of what I was going to talk about has been covered very nicely this morning. So, I think what I am going to try to do is give you some examples of some of the challenges that we have had. And I am probably fairly loud. Am I too loud? Okay, good, because I tend to be loud.

WC#2: **O**f the challenges addressed today,. one of the big things, that patients with underlying hepatitis C don't have specific definitions for application for drug-induced liver injury. Hy's Law was not made for these patients.

Although it has been used as a screening for evaluation and patients at risk, obviously, for potential drug-induced liver injury, it was not intended for that.

We have a challenge with evaluation of the liver biochemistries. We know now, with these very potent direct-acting antiviral therapies, the DAAs, that, in general, all patients get improvement of their liver biochemistries once they start these therapies. So, they start at an elevated baseline. They come down nicely to a normal value, and then, how should we evaluate them? And this has been addressed today. Should we be evaluating them from a baseline, from the nadir on treatment, or should we be using that standard times upper limit of normal, which is usually used in most protocols at this time?

Also, we know that presentations, clinical presentations, can vary quite a bit, and it can vary of lots of different factors, the comorbidities, concomitant medications, the stage of disease, and, also, potentially, genetic factors.

So the FDA DILI guidance doesn't have a lot of information about what you do with patients with chronic hepatitis C. In fact, it is planned to be updated, as we talked about today. We have also touched on what is the type of injury and how these things may affect what you see and what the clinical management would be.

WC#3: And then, also, what discontinuations and followup criteria are appropriate for different hepatitis C patient populations? What cutoffs should be used? What should be used in protocols? What should be used in clinical management? And the balance between safety and not discontinuing too early because of adaptation or other issues that are ongoing, and where that is, we don't want to have loss of efficacy and development of resistance as well, which is another complication with a viral disease.

Currently, the therapies are basically multiple investigational drugs often being used together. This poses another challenge that is important to think about. Because you are having

several unapproved products within a regimen, it makes attribution for a particular product more difficult. So, you have to look at totality of data. You have to consider the class. We take lessons learned, for example, from the HIV realm, where protease inhibitors, for example, have been known to have a risk of hepatotoxicity.

WC#4: And then, as I already talked about, the different host-factors and, also, an immune response to clearing of the virus, you know, when we are starting to see these new DAA regimens used in combination without interferon products, and is there some difference in certain patients, in certain populations where an immune response may be responsible for some of the injuries?

WC#5: This is an example of a published case that came out in Hepatology in 2015 in January. This is daclatasvir and asunaprevir. So, daclatasvir is an NS5A inhibitor used in combination with asunaprevir, an NS3/4A protease inhibitor. These drugs are approved in Japan currently for treatment of chronic hepatitis C.

They are not approved in the United States at this point.

WC#6: This case report, the details are there that you can look at, but it is a little bit easier to look at this in the graphic representation here.

WC#7: So, this is a 57-year-old male who had genotype 1B and was started on the combination of daclatasvir and asunaprevir for a planned course of 24 weeks. As you can see, in the dark line, the black line is the ALT trend, and the red line is the eosinophil. You can see that he started with a normal eosinophil count and a slightly-elevated ALT at around 100. It came down nicely when he started therapy with daclatasvir and asunaprevir. By week two, he had clearance of or improvement in his HCV-RNA level. And he developed fever and a rise in his eosinophil count.

At about week four, he was re-seen and had an ALT up to about 600 and a fever still, with a significant rise in eosinophils. At that time, drugs were stopped and the patient had a liver biopsy. That revealed focal lobular necrosis with

inflammatory infiltrates of eosinophils, lymphocytes, and plasma cells, and hepatic lobules in portal areas. And he also had interface hepatitis and some bridging fibrosis.

The therapy was stopped and he was started on a prednisone course, represented in purple. It was tapered over time, and he nicely responded to prednisone with resolution of ALT abnormalities. Unfortunately, this patient did not have a virologic success and did develop some resistance-associated polymorphisms.

WC#8: So, the authors of this article state the overall clinical syndrome was typical of a drug fever or a drug hypersensitivity syndrome rather than DRESS. There was not a rash component with this product or this case.

In the Japanese trials for daclatasvir and asunaprevir, 16 percent of the subjects had ALT elevations and 9 percent of those had ALTs five times above upper limit of normal. Now this syndrome as well as the ALT and fever appeared to be more frequent in the Japanese patients when

compared to the U.S. and EU counterparts. The author suggests that a genetic basis may be prevalent for these liver findings.

WC#9: Another example is the impact of the stage of disease. So, it is unknown how patients with more advanced liver disease may respond to many of these regimens. We don't have large safety databases with advanced cirrhotic subjects most frequently at approval. And these patients are also the ones most in need of urgent treatment, and this leads often to use for compassionate reasons, which is understandable.

WC#10: In fact, for the case of sofosbuvir, which is an NS5B polymerase inhibitor, and simeprevir, which is a protease inhibitor, this combination was actually recommended in the treatment guidelines prior to approval and is now approved as a combination in November of 2014. Basically, it was often used in patients with advanced disease. Simeprevir itself is labeled as not recommended for patients with severe hepatic impairment.

WC#11: Okay. So, has been published online. This is two cases of hepatic decompensation using the combination of sofosbuvir/simeprevir as a compassionate use for these patients.

Now both of these patients developed marked hyperbilirubinemia out of proportion to their aminotransferases elevations, despite clearance of their HCV-RNA. And the authors' point is that it could be due to the impaired metabolism or underexpression of specific hepatic transporters, and they state that the protease inhibitor simeprevir should be used with great caution, if at all, in patients with more advanced disease.

WC#12: Another example of some challenges that we have had is with another recently-approved product, Viekira Pak. This is a co-packaged and fixed-dose combination of ombitasvir, which is the NS5A inhibitor, paritaprevir, which is the protease inhibitor, along with ritonavir, which is used as a booster. And then, it is co-packaged with dasabuvir, which is the NS5B-palm polymerase inhibitor.

WC#13: So, some of the factors that have complicated evaluations of potential DILI with this product are within a healthy volunteer trial. For drug/drug interaction with estrogen-containing oral contraceptives there was a noticeable increase in ALTs. At risk of elevation in ALTs in females using systemic estrogens was also seen in the Phase 3 trials with a percent, about 9 percent incidence over 1 percent for the overall population.

Paritaprevir is a known inhibitor as well of the bilirubin transporter OATP1B1. That led to asymptomatic elevations of predominantly-indirect bilirubin levels. This is also complicated by the fact that patients are usually using Viekira Pak in combination with ribavirin, and ribavirin causes a hemolytic anemia that also increases indirect bilirubin levels. So, it gets very difficult to sometimes tease out all these variables in particular patients across clinical trials and ascertain the etiology of potential drug-induced liver injury.

WC#14: In summary, these confounding issues that we have gone over today, you know, the drugs that we are talking about, the DAAs, they do concentrate and are also metabolized in the liver, have various transporter effects. And so, drug/drug interactions are an issue.

Patients have various stages of disease and different presentations. Genetic factors could play a role in particular drugs or classes, and class effects may be important.

WC#15: So, there is a lot of commonality between what FDA reviewers are grappling with and what has been presented today. We are in the same boat.

How often are we supposed to monitor these patients, not only really from what is reasonable in the clinical trial, but also what is reasonable or what clinicians will do once we make some sort of recommendation potentially in labeling or what should happen down the road?

What levels of change warrant modifications to a monitoring plan and/or discontinuation?

Again, that careful balance between safety and not losing efficacy.

And what values should be used? Are we talking about increases in baseline, from nadir values, upper limit of normal? Are there particular patients that are at more risk and that need enhanced monitoring or should avoid certain drugs and classes? I know that was part of the talk about the biomarkers as well. And then, could other factors such as race or host or viral factors contribute as well?

WC#16: All right. Thank you. (Applause.)

Session IIA Discussion

DR. SENIOR: Thank you, Wendy. I think you can tell how fortunate we are at the FDA to have such excellent medical reviewers, clinical reviewers. Both Wendy and Karen are examples of the quality of people who are looking at the data from clinical trials. They are really doing a first-rate, outstanding job. Thank you very much. (Applause.)

While people are going to the microphones for questions, I want to put a question to Dr. Chen. Please come to the microphones for the discussion. But let me ask Dr. Chen: I thought I heard you say that you were going to modify the recommendations for what actions should be taken for the different grades of abnormality of tests. I thought I heard you say that you don't die from an elevated transaminase, and so forth; you die from the disease that is causing it. What I am saying is, shouldn't action be taken when you get a high value for a liver test means that you have high imperative for taking action to find out what is going on, to investigate? In other words,

it is not a measure of severity, but a measure of urgency to discover what is really going on.

DR. CHEN: The CTCAE is a grading document. It is not a recommendation for further evaluations. In the protocol itself it can certainly state that patients who have Grade 1 or Grade 2 adverse events require further investigation, but I think that is what we do in general. If there are any abnormal labs, the practice is to try to figure out the cause of it.

But I guess the question for this group is, do we need to go all the way out to Grade 3 or 4 for AST or ALT? But, in terms of what actions that need to occur for certain grades, I think those are more determined by the protocols than by than the CTCAE.

DR. TEMPLE: John, were you making the point that you shouldn't necessarily stop the therapy from that?

DR. SENIOR: Not necessarily.

DR. TEMPLE: Or what did you want to communicate here?

DR. SENIOR: Well, I think we need to argue this, debate it, and come to a consensus. We need to face the issue of what are we talking about. Does an elevated transaminase that is 20 times upper limit of normal, whatever that is, mean that you are in imminent danger of death? No.

DR. TEMPLE: Well, that's right. We have learned that over and over again. There are some drugs that do that regularly and somehow never lead to major problems.

DR. BJORNSSON: Thank you. I have a question for Dr. Carter.

DR. SENIOR: Everybody may know you, but tell everybody who you are.

DR. BJORNSSON: My name is Einar Bjornsson. I am originally from Iceland, but I have been at the NIH now for nine months. I have a question for Dr. Carter: I wonder if you are aware of any studies on how often patients with chronic hepatitis C have a significant elevation in liver enzymes? Let's say ALT more than 500 spontaneously.

DR. CARTER: So, flares, basically, yes.

DR. BJORNSSON: Yes, yes, flares. I know they are rare, but how often?

DR. CARTER: Right. I don't know that there is an exact number from the data, but I know there are papers out there regarding hepatic flare and that that can occur in hepatitis C, although it is described much more obviously in hepatitis B. But I am not aware of the specific number. If there are others, somebody else, that would be great. Chime-in.

DR. DUNN: Again, I am Laura Dunn with the FDA. Well, I can't answer that question either, but that actually ties into my statement and appeal. We are trying to update the DILI guidance to deal with patients with abnormal baseline liver functions. This is enormously difficult to figure out. One of the things we don't have is good data on baseline variability, the normal baseline variability in different subsets, hepatitis B, hepatitis C, cirrhosis, NASH. I am actually working with Dr. Sanial now to query the NASH CRM database

to give us some data for NASH. I do have some datasets in-house from some cirrhosis populations. But we really need to gather together the data on these populations, because we don't know how much change, you know, depending on where you start, how bad you are when you start, how much change is normal, kind of within the normal variability and how much change means we should be concerned and do a workup.

So, my plea is that we need more data, possibly through the Liver Forum, which isn't really a DILI forum, but we are getting sponsors together and trying to work on sharing data, especially maybe placebo group data. We might be able to work on that.

DR. SENIOR: May I ask what data? We want the data on the people who start out relatively normal, to give us insight as to whether an increase in someone who starts out abnormal is a problem or not? If the normal people never go up, if there is no evidence of transaminitis in them, there is probably more reason to think that someone who

starts out high is just bouncing around, but probably not related to the drug, right?

DR. DUNN: Well, I mean, when you are trying to like write a protocol and say, "When do I check these patients" -- You know, when? Is ALT two times baseline? Is that concerning in this population or is that normal? I mean, that has been a struggle for us. Mark and I have had discussions internally about if you start between two and five times baseline, you probably need a different range than somebody who starts at 10 times baseline. You know, you need different parameters.

DR. SENIOR: Okay. Leonard has reached the microphone. (Laughter.)

DR. SEEFF: If I could make a comment about it, when I was with John, we were concerned about this. What is the variation in the ALT in people with hepatitis C? And we turned to Harvey Alter at the National Institutes of Health. Do you know Dr. Alter? Dr. Alter is one of the most famous people in the study of chronic hepatitis C at the NIH Blood

Bank. He had collected over the years patients with chronic hepatitis C entered into a trial and he had been screening it on a regular basis over years. We actually have been trying to work with him to do a study. John and I were hoping that -- he has started with us, but we have not gotten around to completing this. The impression I got from him is that there is not a great deal of variation in the abnormalities. It is simply true that, the lower the abnormalities, the less likely there would be variation. The higher the abnormalities to begin with, the more likely that it is to be variation. I don't have a number on that. I don't know, John, whether you have anything new to add to that. But it is that kind of data that I think would be very helpful.

He has an enormous collection of data on this particular aspect. I think this is very important because let's assume the enzymes are running at 100 plus or minus, and suddenly it is 500. Is that to be expected in a person who has chronic hepatitis C? Or is this a signal of something more

important? So, this was, of course, the issue that we had.

DR. SENIOR: Good question. Okay.

DR. DUNN: When you are using a drug that is treating a liver disease like hepatitis C, I think you get an improvement in your liver functions. And then, you have a new nadir. And so, when that changes, you know, that is a pretty good signal.

But when you are using a drug, say rifaximin, and you don't expect it to affect your baseline liver functions, for that population when there is a change, it is a little harder.

DR. NORRY: Elliott Norry. I am a clinical safety physician at GSK. I have a question for Dr. Chen. With respect to the CTCAE criteria, I work quite a bit in the realm of drug-induced liver injury in cancer patients. I am wondering there is thought to adding the AE term of just drug-induced liver injury to the CTCAE terms.

And let me qualify my question because we see a lot of Grade 2 ALT abnormalities, a lot of Grade 2 bilirubin abnormalities. When you have a Grade

2 ALT abnormality along with a Grade 2 bilirubin abnormality, it is really not sort of Grade 2 anymore, if you are going to get sort of past just the laboratory phenomenon.

And the only hepatic or hepatobiliary term available is hepatic failure, which implies that the patient has encephalopathy and is only available at Grade 4 and Grade 5. There is a whole lot in between Grade 2 laboratory abnormality and Grade 4/5 hepatic failure that I think would be worthwhile to capture.

DR. HICKS: No, that is part of the reason for being here and participating and getting the experts here to help in terms of one question is, do you know if DILI is a MedDRA term?

DR. NORRY: It is a MedDRA term, but it is not a CTCAE term.

DR. HICKS: No. So, if it is a MedDRA term, then we can certainly consider it. In terms of grading, we would appreciate any recommendations.

DR. NORRY: Yes, I think that your grading of 1 through 5 probably could be applied just based

on the general guidelines of how you grade 1 through 5. But, currently, I see cases that are sort of interpreted as Grade 2 that could have Hy's Law with ALT at five times the upper limit of normal and a bilirubin of three times the upper limit of normal, and they are just sort of captured as a laboratory abnormality within the CTCAE terms.

DR. HICKS: Thank you.

DR. REGEV: Arie Regev from Lilly. I have a question for Wendy Carter regarding one of the articles that you showed on the slide. I didn't read this case before, but I tip my hat to the people who made the diagnosis of DILI in a hepatitis C patient that had jaundice, and they decided it was related to the drug. I was wondering if people that were involved in this decision are sitting here, and could they share a little bit how they decided this was drug-induced?

DR. CARTER: Well, I'm not one of the authors of that article, but I don't know if anyone is.

DR. LEWIS: I am. Jim Lewis, Georgetown.

DR. CARTER: Awesome.

DR. LEWIS: These are two cases that were at the University of Virginia on the transplant list. They were stable on the transplant list. They were treated in a protocol pre-transplant with Simsoft, the sofosbuvir and simeprevir combination, and they both completely deteriorated on treatment. They had been stable. We did a RUCAM analysis. It came out at 9, which was probable DILI. And the general impression of the people taking care of them, the transplant hepatologist and the others, we thought it was important enough because we were worried about the protease inhibitor part of this causing a problem. As you said, it is labeled that it shouldn't be used in severe liver disease, but it was. And so, it was as close as we could get to figuring out that this was due to the drug and not due to some change in their underlying disease. So, that is why we said it was probable. You know, it was probable; it is not a perfect assessment, but it is as close as we could get.

It has engendered some other articles as well. There is another report on simeprevir that is out there.

DR. CARTER: Absolutely.

DR. LEWIS: And you mentioned the case with asunaprevir, which is the problem. It is not daclatasvir. It was the asunaprevir. At least that is my impression.

DR. CARTER: Yes. I appreciate your feedback. That's great, yes.

DR. SENIOR: There is a clue. What we have in the cases of viral hepatitis is we have some extra information that is always gathered. It is called the viral load. It is the number of particles of virus that are in the circulating blood, in the plasma. And you actually count them. Now, usually, the treatment reduces the count of the viral particles in the circulation from, say, 10,000 down to undetectable. Undetectable is around 50 or something like that.

If the injury to the liver is occurring while they have completely suppressed the virus, then you

want to say, well, maybe it is not due to recurrence of the disease, recurrence of the viral hepatitis, but might be due to something else; namely, the drug. So, it is a difficult point of information. Ted Guo and I are planning to incorporate the viral load data into eDISH-2 for cases of treating viral hepatitis with these new drugs.

DR. TILLMANN: Hans Tillmann, ECU. Two comments. One is I think for HCV it always occurs that when the virus drops, the transaminases drops. For HBV, it is a little bit more complex. And now, the field is moving to hepatitis B. I would not be surprised if we see that people are completely suppressed and, then, get a flare which is actually due to the immune reconstitution and, then, helping to finally eradicate hepatitis B.

While I would agree with what you have seen, Dr. Lewis, in simeprevir patients, when you suddenly see a flare in hepatitis C, you would be concerned that that is not virus and immune-related. In hepatitis B it would be difficult.

DR. CARTER: I agree.

DR. TILLMANN: Or it would be likely to be different. We don't know yet.

DR. CARTER: Right, and I think that is just the nature of the differences. And then, also, you have to feed into what you know about the investigational agents. Like we say, for example, protease inhibitors, it is not really too surprising that there may be some issues regarding that. But it is definitely part of the disease and, then, the host as well.

DR. TILLMANN: Can I comment on the fluctuation of liver enzymes? I have now seen a number of people who show up with normal liver enzymes and four times the upper limit of normal, a few months later normal, elevated again, without identifying the reason. And I have even seen healthy people coming up with liver enzymes in the thousands, which then normalized, which certainly makes the job to identify is it really drug-induced even more difficult, that probably sometimes in the background we have fluctuations.

DR. DUFOUR: Bob Dufour from the VA. The VA has a very large national database on hepatitis C that might be a good source. We had actually published an abstract about 15 years ago looking at the correlation between ALT levels and histologic activity in liver biopsy samples. We found that those people whose ALT levels did fluctuate tended to have more severe liver injury. So, it does correlate with that. And I could probably go back and pull that data out and look at what degree of fluctuation we were seeing. But there is more data that exists on a national basis.

DR. CARTER: Do you know the percentage of how many were fluctuating versus being stable?

DR. DUFOUR: Again, this was 15 years ago.

DR. CARTER: You don't have that right there?

DR. DUFOUR: I don't have it with me, but we do have that data. So, I will get it to you.

DR. CARTER: That would be great.

DR. DUFOUR: Communicate with me. I can get you that.

DR. CARTER: That would be great data to have, yes.

DR. TAWAZAM: Hi. Just one question. This is Quay Tawazam from Derry. I just have a quick question. Are there any data showing specific genes or genetic predispositions of one group who are more predisposed to drug-induced liver injury? Because I saw the article, I believe, in your presentation in Hepatology. They were proposing a genetic predisposition. But I haven't seen any other data. And also, the background to my question is a lot of the Asian population you do see hepatitis B/hepatitis C infections which are more prevalent. And then, you have these treatments that also affect your enzyme level. Then, if there is also a genetic predisposition, how do you even tease that out, whether it is the drug, it is the disease, or some other component? So, I was just wondering if there is any --

DR. CARTER: That is exactly the challenge. Right now, there is no published data. And it may be a combination of the genetic predisposition and

the particular products versus other issues, whether it is just genetic predisposition across the board. But, as far as I am aware, right now there is not anything identified and specific to that Japanese prevalence as well.

DR. TAWAZAM: Right, yes. Also, one of the reasons I am asking, for some products we have we do see a lot -- we don't see any data in our clinical trials, but in our post-marketing data we see a lot of reports where some of the liver toxicities are all coming from a certain region in the world. So, that is why I am curious if there is any specific genes.

DR. TEMPLE: That is of major interest. It would have to be a very important drug for anybody to want to go through the trouble of typing everybody. But I saw data on lumiracoxib that suggested that almost all the Hy's Law cases were in people with a particular genetic characteristic, and it might have looked somewhat regional, too. That should be an area of major interest, at least for drugs we care about. There

have been metabolic differences in the past that predicted problems, but I don't mean that.

DR. TAWAZAM: And especially in our population, you also have like over-the-counter herbal supplements which usually you don't have -- I mean, you are not collecting them in the clinical trials. So, it is hard to tease them out.

DR. CARTER: That is asked about, though, routinely. I mean, at least in our clinical protocols, especially if someone has a risk. That is something that we do specifically ask for and do follow up with with investigators, if we are aware of cases, because it is very important to find out if there are other over-the-counter herbals. We do try to make an effort to do that, and, as well, I am sure the industry folks will agree that is something that we have had many discussions about being thorough in the review.

DR. AVIGAN: Just a little followup. This is Mark Avigan, at FDA. So far, the experience of genomic markers and susceptibility to liver injury is really drug by drug. It is a little bit all over

the place. With lumiracoxib, that was kind of very dramatic, that there was a particular marker, a Class 2 marker, which is invariably for the very severe patients highly enriched for risk.

Some of the other drugs, in the case of simeprevir, for example, which was shared, that did not seem to have at least an HLA isomarker. In particular, there are examples in the Asian population with Carbamazepine, for example, and the marker B1502, which, again, very important and is in the drug label --

DR. TEMPLE: Right. Not for liver; that was for --

DR. AVIGAN: That is a hypersensitivity reaction, but that is kind of an example. But, for liver, there really has been, for most of the markers, there have been very small risk effects, not large risk effects. I wanted to actually just follow up on the point Laura made, where we really are struggling with patients who have significant liver, chronic liver disease, where they are now being introduced into protocols either as the

target population for treatment, more of these kinds of patients, and not just one liver disease, but different liver diseases, NASH, for example, as well as chronic viral hepatitis, and patients now with cirrhosis who are being treated to try to mitigate the progression of their disease.

When we see worsening of liver injury where there is already a high MELD score, for example, and what the protocol rules should be with regard to how you, then, ramp up monitoring or stop rules, for example, becomes very problematic.

So, if you start with a patient -- let's say your target population are patients who have a MELD score of, let's say, 5 or even 10, where we have seen such cases, how far would you allow worsening to go before you really got worried about the drug rather than the course of the disease itself?

And so, these are very challenging questions and ones that we would like this group to start grappling with, as we are working on this aspect, a very effective guidance, but one that perhaps

does not address this question of underlying chronic disease.

DR. TEMPLE: Mark, on the same question I had before, wouldn't a lot of what you determine to do have to do with what the drug was doing in normals? I mean, if it never caused transaminase elevation in normals --

DR. AVIGAN: But there is a lot of nuance. So, what you now have is a perturbation, not just because the liver is the site of drug metabolism, but also with the liver there is the potential effect of shunting and secondary physiologic aberrations because of the liver disease that changed the way the drug works on the patient.

So, for example, if the drug were -- if there was shunting as a consequence of cirrhosis and the drug did not go in first-pass metabolism to the hepatocytes, but went somewhere else, that could become a consideration that the drug was toxic. That is a pharmacodynamic or pharmacokinetic effect.

But, in addition, there could be a natural course of the underlying liver disease where you were using the drug for compassionate use, as we heard that more and more of these drugs are being used kind of off-label for these more severe patients. And then, there was worsening in that patient population. How would you dissect out the effect of drug as a toxicity effect versus the natural course effect?

DR. TEMPLE: The drug isn't particularly a hepatotoxin, but it does something to the underlying systems that make the disease get worse.

DR. AVIGAN: It is a possibility. And again, I think from the point of view of drug development, what we need are clear guidelines or rules from the point of view of how to establish protocols that allow us to identify, to characterize the risk, but also to protect patients in studies, so that they will not have an untowards safety effect where they have less buffer before they get liver failure.

DR. WATKINS: Just to follow up on that, I agree with Bob. I think Hy Zimmerman said, if you are talking about idiosyncratic hepatotoxicity, it is no more likely to occur in the individual with preexisting liver disease than the healthy liver. That was his observation, and I don't know anything that contradicts that.

DR. DUNN: Yes, but no more less likely, either.

DR. WATKINS: Well, that's correct.

DR. DUNN: So, you still have to address it

DR. WATKINS: No, no --

DR. DUNN: Because when these drugs are in trials, we a lot of times don't have that much information. It is not like we have hundreds of thousands of patients with normal livers who have taken the drug. So, I may not have a really clear idea --

DR. TEMPLE: No, but you do have enough people to know whether the drug regularly raises transaminase. That is very common.

DR. DUNN: Right, but if I don't know of a major signal, I still need to say, okay, I don't have a major signal, but I still need rules for DILI in this protocol.

DR. TEMPLE: Sure you do.

DR. DUNN: So, how could they be written? And that's --

DR. TEMPLE: But what I am asking and what Paul is saying is, if the drug shows no indication of being an injury drug, no transaminitis at all, and then, we see somebody with preexisting disease who gets worse, I think your going-in bias is that it is the disease fluctuating now. You know, what you are going to do to protect the patient is another matter. But you do have other information on whether the drug is actually toxic to the liver cells.

DR. WATKINS: I mean, I agree completely we need guidance on this. But I am just saying, if you have a particular combination for hepatitis C and you move into the decompensated cirrhotic, the first question is, what did you see in the

compensated cirrhotic? And if you saw nothing there, it is less likely that there is an issue you are looking at.

And also, drugs tend to have a signature in terms of latency. If you see decompensation in that same latency as you saw ALT elevations in more compensated liver disease, it makes it more likely. But I completely agree we need guidelines.

DR. DUNN: Yes, but, I mean, it is not that easy in decompensated liver disease throughout a population, either. And then, what is a normal variation based on?

DR. WATKINS: Sure.

DR. DUNN: I mean, you are saying that all these drugs are going to have data, a lot of data for patients with normal livers, but when these drugs go to trial there's not a lot of data.

DR. QAZI: One more thing is that these patients are on multiple medications. So, that adds to the conundrum.

DR. CARTER: Well, and the other issue is that sometimes patients do get elevations of their

liver enzymes, and they are able to stay on therapy and continue and they have adaptation, or whatever you want to call it, or resolution. And so, it is hard, even when you see that. Is that the signal or is that not the signal?

DR. TILLMANN: Hans Tillmann, ECU. It could also be that in an advanced liver disease you might have a mechanism which usually protects you from toxicity which is not as active anymore. And therefore, it might be that you only see it in cirrhotics. So, even though probably one would start with fibrosis, and some of this fibrosis goes to advanced or to compensated cirrhotics, and only after, then, to the decompensated cirrhotics, it might be, well, that you only get a safety signal in the decompensated cirrhotics.

DR. WATKINS: Theoretically, but give us an example.

DR. TILLMANN: I do not have an example, but I think we need to be prepared –

DR. LEWIS: Jim Lewis again from Georgetown.

First of all, the compensated cirrhotics, it is finite period of time they are going to be treated. Usually, it is 12 weeks, maybe 24 weeks. But, with some of the protocols for the patients on the transplant list, all of whom have MELD scores that are going to -- to get on the transplant list, you have got to have a MELD score of at least 14 or 15. And the shoulder of the MELD score starts falling off around 20.

So, all these people who are being treated, we are trying to prevent a transplant. They are all high MELD scores. And so, it is a different kettle of fish a little bit. These people were not studied in the clinical trials, you're absolutely right. We may have to make special rules for them. But, at the moment, it is on a case-by-case basis, how they are doing. We try to keep them safe.

Obviously, it is better if we could cure them and they don't need a transplant. One of the succeeded. The other one did not; he needed a transplant. But these are the sickest of the sick people, and there's a lot of things going on. Can

we really be sure it was the drug? But the people who take care of these folks have a pretty good understanding of what the underlying disease is doing. If the virus is controlled and they are getting better, and then, something happens, then we have to look for either the drug did something in this particular patient that we wouldn't normally expect, because normally we wouldn't expect worsening liver disease in the compensated patient.

I don't know that we have a lot of information on people with completely normal livers. Maybe in the Phase 1 there were some, but I don't think there was any problem. I mean, almost all these trials now are done in patients with the disease because you can't withhold the treatment anymore because it is so good. But, yes, I mean, if we are talking about fold elevations and ALT and stuff, I think MELD scores probably are a reasonable consideration for stopping, you know, when we start writing additional guidance or something. But the MELD score would have to be quite high.

DR. DUNN: I think your point from your clinical perspective was that you are doing very short things in transplant patients. But I'm looking at drug trials of six months or longer in the drugs.

DR. LEWIS: Oh, no. No, in the transplant patients, these are patients who may be on the drug chronically and certainly post-transplant. When we are treating patients who have recurrent hepatitis C post-transplant, we don't know what the end-game is there in terms of the duration of treatment. There are protocols right now that are looking at that, whether it is going to be six months or a year, so that we can keep these people from having to have another transplant.

DR. DUNN: All right. Again, I think clinicians do have a very good feel when they are taking care of an individual patient as far as when to stop the drug, but nobody knows how to write it into a clinical protocol. That is the problem.

DR. LEWIS: I will just tell you at the University of Virginia they will no longer use

protease inhibitors in anybody on the transplant list. We had to write that out of the paper. The editors didn't want us to say it that strongly. But I will tell you that that is how they feel down there. They will not use it. And we are way away from that particular combination now. That is no longer even being used.

DR. TEMPLE: We are very close to needing to take a break. Anything urgent?

PARTICIPANT: No, I just have one question. I think you also learn from exceptions in science. So, I was looking at your table of the twofold and the low P-value and the clear exception of diphenhydramine, and the use more to protect CNS activity. And what the combined intelligence in the room had, that that lesson shows that it is just out there and not liver toxic. So, what is it that that is telling us? Thank you.

DR. TEMPLE: Well, I think I am going to raise this a little bit later. How to figure out which drugs that give transaminase elevations are really not going to cause problems is one of the major

tasks. We know there are some, but to earmark them or identify them and really know is not so easy. I don't think there is a current mechanism. Sort of experience tells you. Last question.

DR. QAZI: Well, I broadly agree with you, Jim, in terms of using MELD as a marker. But the thing is, one has to be aware of the nuances of the MELD. One of the competences of the MELD is bilirubin and it is impacted by hemolysis or like a stone sitting in the bile duct. As long as people are aware of that, because I have seen people call -- they're writing DILI event -- grand hemolysis DILI.That's it.

DR. TEMPLE: Okay. Break time. When should people come back? Three o'clock.

DR. SENIOR: 3:05.

DR. TEMPLE: OK, 3:05.

(Applause.) 2:46 pm

Refreshment break

Session IIB

So, we only have Dr. Temple for this afternoon, but I am sure that he will provide us -- sorry, Bob

(laughter) -- I'm sure that he will provide us something that will engage us in stimulating conversation.

DR. SENIOR: I wanted to say something about Gaby Danan.

MS. PAULS: Oh, go ahead.

DR. SENIOR: I'm going to talk a couple of minutes. Dr. Danan sent an email yesterday -- it came in at eight minutes of 1 pm, which is 7 o'clock Paris time -- saying that his plane to Iceland, to Reykjavik, was going to connect to a flight to Washington. He was going to fly in late last night. He sat on the ground in Paris for four-and-a-half hours and was too late to catch the connecting flight and too late to get here today.

Now I want to say a couple of words about Dr. Danan. He called back and spoke with Lana on the telephone. And he said, "I don't want anybody to show my slides." (Laughter.) Okay. Well, I have his slides and I will be glad to show them. But he said, "No."

But let me tell you a little bit about him. Gaby Danan, along with his mentor, the late Dr. Christian Benichou, while working at the company called Roussel Uclaf in Paris had gathered together a group of French hepatologists in the mid-1980s, 1985 to 1989. They worked together and they talked about the problem of liver injury caused by drugs. This was in France. And really, the problem wasn't recognized by hardly anybody at the time. And they convened in June of 1989, 26 years ago, a group of 12 distinguished hepatologists. There were five from France: Danan and Benichou, plus J.P. Benamou from Clichy, Bernard Begaud from Bordeaux, and a Dr. Legiere from Paris. There were two people from Switzerland from the Council of International Organizations for Medical Sciences, CIOMS, who were there for their political clout. There were also two from the United States, Hy Zimmerman and Willis Maddrey. And there were one each from Denmark, Nils Tygstrup; one from Italy, a Dr. Orlandi from Anconia, and one from England, Neuberger from Birmingham. Twelve people. Of the

12, only two or three are surviving, including Drs. Danan and Maddrey.

I don't know if Will wants to say a word or two about his recollection of that meeting back in Paris, when they started the whole ball rolling about diagnosing DILI. Willis, any comment?

DR. MADDREY: Understand this is totally staged. John warned me that he might do this early this morning, but I was counting on him forgetting. (Laughter.) We spent so much time last year talking about RUCAM versus other ways to evaluate. And I had rather hoped that the discussion at least would now be laid to rest. But, since it is coming back up, I want to give you just a recollection or two about that meeting.

That meeting was a terribly serious event. Everybody came prepared and spent a great deal of time working on this system for causality assessment. And it is remarkable that an idea has lasted this long in modern medicine, and many of the tenets that are certainly the basis of what we do have changed. What I remember about the whole

thing is that Zimmerman and I had a marvelous weekend. And I can remember the menus better than I can the discussions. (Laughter.)

But, as I mentioned to some of you in the past, the best thing was one of the few things in my career that I am proud of is that I was able to negotiate Concorde tickets for Zimmerman and me. (Laughter.) I was considerably younger than Zimmerman. And I convinced them that his health was such that we really should fly, and they couldn't have the meeting without him and he couldn't come without me to take care of him. (Laughter). So, we got to fly on the Concorde, on Air France. And Dr. Zimmerman got a chance to meet some beautiful young women. They chose flight attendants in those days quite differently than they do now. (Laughter) It is nice for me now to get on planes and I meet some grandmothers whom I have known along the way. (Laughter) But in those days it was different, and I talked to these young ladies and said, "Please make Dr. Zimmerman happy," and they did. And he was extraordinarily happy.

I had a hard time getting him out of France. (Laughter.)

But I would say about the RUCAM that we had such a discussion about it last year. There are many good points to it. Those of us who are on the side of expert opinion really use all the points that are in RUCAM. The only things that have come to the fore are really nicely capsulized, if any of you want to read it, in the program run by Jay, the LiverTox Program. The very first part of the LiverTox Program has a three or four-page learned discussion of RUCAM, and the pluses and minuses of each of the components of RUCAM and how they have changed over the years. So, John wanted me to mention this. Gaby is a wonderful fellow, and Dr. Benichou was a driving force in all this. But all I can really remember was the great meals, and there was this one woman on that plane. (Laughter.)

DR. SENIOR: Thank you, Willis. Well, that is just fascinating because -- (laughter) -- RUCAM is still alive and well. I think a lot of people are still using it as opposed to what the DILIN Network

says. They have rejected the use of RUCAM as a scoring method to diagnose DILI in favor of expert consensus. We are experts, but not everybody else is. So, I understand that the DILIN Network, now in its I think 13th year, is not going to be renewed for a third five-year cycle, according to the rumors from NIH. So, I think there is a challenge here that the DILIN Network of experts doesn't go out of business before they give us something better than RUCAM.

Now I would like to take a rough poll here. How many in the room still use RUCAM (Show of hands) Well, there're quite a few. How many use expert consensus only? (Show of hands) A lot of people didn't answer. But it looked like about equal.

So, RUCAM is still a player, and it isn't bad. It can be used by ordinary physicians. You don't have to be a hepatologist to use RUCAM. So, it still has its use, but I think before the network of really expert, well-recognized hepatologists in the DILIN Network go out of business, they should

provide something to replace RUCAM for the average physician. So, we need an alternative.

Bob Temple. Bob?

Temple photo, links to biosketch and abstract

RT#1: Okay. Well, I was assigned the task of talking about -- this is all John's fault, I assume -- limits of labeling and warnings. So, I am not really going to be talking about most of the stuff we have been talking about. I am going to try to talk about what we do when we get certain information. And I am going to do it in several cases, just to lay out what we generally do. But, as I went through this, I realized that, although we know which drugs have been rejected because of liver toxicity, which drugs have been withdrawn because of liver toxicity, I am not sure what we have done with all the ones where there is a little suggestion of toxicity and we thought there was some problem, and I can't tell you whether we are enthusiastic about monitoring all the time or what we say, because I don't think we have taken

a systematic look. But I looked at a few of them. So, I will tell you what I looked at.

RT#2: This is in part how we apply risk/benefit considerations to evidence of liver injury. So, take the first and, if you like, clearest case. This is going to be mostly about what labeling would say regarding liver injury and monitoring, but that depends on what the drug is for, what the alternatives are, and the nature of the injury.

So, take Case 1. You have clear Hy's Law cases, say at least two. We have done this sometimes when there was only one, I have to admit. Or, in the post-marketing period we found unequivocal severe hepatic injury. We may have missed it during the pre-marketing period or we didn't interpret it right, or whatever. And there is plenty of available therapy, either pharmacologically similar, which is sort of easy, or even mechanistically distinct, and just no documented advantage over alternatives. That doesn't mean a new method of working couldn't prove

to have an advantage, but at least at the moment it hasn't.

I would say the regulatory conclusion, if we recognize a Hy's Law case, is invariably non-approval. That is what we did with ximelagatran and lumiracoxib and dilevalol. I always like to mention those because they were all approved in Europe and subsequently withdrawn for hepatotoxicity. So, without being smug, I am just taking note of that. (Laughter.)

Two very similar cases were the withdrawal of bromfenac and troglitazone, which conceivably could have been rejected in the first place on the basis of Hy's Law cases that were, in fact, present and abundant evidence of transaminitis. What we tried to do with bromfenac was to limit it to short-term use, because most of the cases of problems occurred after a while. In retrospect, that was an implausible thing to do for a nonsteroidal anti-inflammatory drug, which is plainly intended for long-term use.

And troglitazone, interestingly, was left on the market after its hepatotoxicity was unequivocally discovered, with a request for monitoring which plainly failed, because it was a unique anti-diabetic drug. There was nothing similar to it. So, we withdrew it only after we watched the two follow-on drugs, ROSI and pioglitazone, and satisfied ourselves that they were not hepatotoxic, which took about six to nine months. We had a working group that met every month or few weeks to see if they looked clean, and then, we yanked it.

Probably, if I had been signing off on those, I wouldn't have signed them because of the Hy's Law. But, anyway, we have it now, and it is quite clear that, if the drug has no advantage and has Hy's Law cases, it's gone. Sometimes that would be true even for drugs that had attractive characteristics. I recall ticrynafen, for which I was responsible party for approving it. We didn't know that it was hepatotoxic before we marketed it, and it was a very attractive drug, a uricosuric diuretic at a time

when one-third of the gout in this country was due to use of diuretics. So, that was not trivial. Of course, that is because we overdosed the diuretics.

It is worth mentioning that in some cases like this where drugs were -- oh, I should say what our expectation is, that if you see a couple of cases in a database of 1,000 people, you are expecting, roughly, 1 in 10,000 or more. This is what Hy would have predicted, 1 in 10,000 deaths or nowadays transplants. Bad enough. Severe liver injury that is life-threatening.

But we have approved other drugs with problems. We approved clozapine with a 1.5 percent rate of agranulocytosis. And how many deaths that causes depends on what you think the survival rate was going to be. We used to think that agran led to about a 10-percent mortality. That hasn't been true with clozapine, but maybe 1 percent. That is in the same neighborhood as 1 in 10,000. But it was approved because they showed that it worked in people who had failed therapy with other drugs. And if you have a drug that treats

psychotics who can't respond to any other drug, that is a big deal.

Similarly, a calcium channel blocker called bepridil, which is a clear QT prolonger. And there are torsade de pointes deaths reported every single year. It remained on the market because they did a study that showed that in non-responders to diltiazem randomized back to diltiazem and bepridil, bepridil was the more effective antianginal drug. So, if you can show some spectacular benefit, you might be able to overcome even Hy's Law cases. And in terrible diseases, oncologic diseases and things like that, we tolerate all kinds of things.

RT#3: Now let's say there clearly are Hy's Law cases, but the drug has worthwhile advantages -- that is sort of what I was talking about -- over alternatives. The drug could be approved or remain marketed with labeling urging monitoring. That is true for isoniazid. And we believe that monitoring reduces the risk of severe liver injury or required monitoring and REMS. And

that is what we did with bosentan, which was the first drug available that was effective for pulmonary hypertension. But the monitoring needs to be realistic. For a serious lifelong illness like pulmonary hypertension, where people come to the doctor every couple of months and stuff like that, maybe monitoring is credible.

It also appears, for reasons that I don't think we know the full answer to, that monitoring seems to limit the likelihood of severe liver injury, because we have seen very few cases of fatal liver injury with bosentan or transplants. I don't know how many, but not many. On the other hand, the call for monitoring didn't seem to do a thing with troglitazone, either because it wasn't done or because it doesn't work. I think probably it is a little of each, but you can go downhill very fast.

RT#4: So, when we would call for monitoring or ask for monitoring, and things like that, it is not terribly well-established. One of the things that I think would be worth considering is, when

might monitoring work? Are the signals of hepatotoxicity different? Is it the steepness of the curve or whether it continues to occur after you withdraw the drug? I mean, I don't know. I have no answer to this. But you would like to know how to distinguish the troglitazone case from the bosentan case, in case you did want to make a drug available with monitoring.

So, the experience to date is not so easy to know. Troglitazone monitoring didn't work at all. Bromfenac monitoring, which they were supposed to do, didn't seem to work. Isoniazid, as I said, seems to work at least some, and bosentan seems to work very, very well. You seem to be able to monitor#5: r your way around it.

So, if there was a way to anticipate this, it could be informative. As we were discussing earlier, if there is some genetic marker that predicts who is going to get into trouble, that would be fabulous, but whether that is even worth thinking about isn't so clear.

RT#5: So, the third case. Aminotransferase elevations, but you really don't have any Hy's Law cases in, say, 1,000 or 1500 patients. And we know that there are drugs, heparin, aspirin, statins, tacrine, very conspicuously caused transaminase elevation, but rarely, if ever, gave you bilirubin elevation or liver failure. And I am not sure we know how to tell which of those it is going to be.

Labeling for these drugs has sometimes called for monitoring. As probably everybody knows, statins did until we decided it was silly because there were never any bad outcomes. And one of the things I realized in getting prepared for this is that I couldn't catalog what we have done with respect to monitoring and calling for liver enzyme monitoring on drugs. I don't think we have ever looked at it systematically. I think it would probably be worth doing to see whether it is really worth it and what we actually get out of it.

RT#6: Now there are also cases where there is severe liver injury, and even fatal, but it is pretty rare. I say "very rare"; I'm not sure what

"very" means. There is no question there are fatal cases of liver injury with labetalol, diclofenac, things like that. But diclofenac is, not in the United States but in the rest of the world, probably the most popular NSAID, and it is not as bad as bromfenac, ibufenac, and something like that. It is probably COX-2-selective, at least a little bit. So, maybe people like it because of the bleeding.

Labetalol also has some advantages. It is actually a diastereoisomer, and it is really two drugs, not one, which we didn't know at the time. It is a beta blocker and it is a vasodilator. It has properties that other beta blockers mostly don't have. So, it is out there, even though we get case reports all the time.

So, those drugs remain out there with warnings. They both called for monitoring, but they give some fatal injuries. And that is either because monitoring doesn't work or it isn't done, and we don't really know. As I said, I would be interested in looking into those cases and trying

to pin them down better than we have to date, which I think would be worthwhile.

RT#7: So, it is pretty clear that serious hepatotoxicity, that is, the kinds of drugs we don't approve in the first place, is not really dealt with by labeling or monitoring. We just don't think that the risk is worth it, and we don't approve them. But, if a drug has an important benefit, like bosentan or isoniazid, we do leave it out and we try to help people get around it by monitoring and stopping, and with the two where we have done that, pretty successfully, I would say.

Diclofenac and labetalol I think need a close look to see just how much toxicity we have and, if possible, to figure out why. Is it because nobody was monitoring or because you can't monitor anyway? The rates of liver injury there are much less than the 1 in 10,000 that we think Hy's Law would predict, but I don't know what they really are. And there are other drugs where this issue has come up. Haldol might have rare cases, and so on.

So, that's all I wanted to talk about. I am interested now and we will try to see if we can do something about looking into what our pattern has been with respect to monitoring and things like that. So, that's it. (Applause.)

Discussion IIB

DR. SENIOR: Don't go away.

DR. TEMPLE: I'm not going away.

DR. SENIOR: There are a couple of things that concern me about the limitations of labeling. One is: do physicians really read the labeling, and even if they read it, do they understand it? The labeling is getting to be 20 or 30 pages long now. Do they really read all that stuff? And third, do they follow it? So, do they read it? Do they understand it? Do they follow it? And the answer is no, no, no. So, isn't that a really significant limitation to labeling? Just labeling it doesn't solve the problem unless it causes action.

DR. TEMPLE: Unless the drug has a really meaningful advantage, that is the reason that we say no if it looks hepatotoxic. Now the ones in the middle are drugs like labetalol and diclofenac, where I'm uncertain why we are not either more worried or don't have stricter limitations.

It was easy for bosentan; you have to come and get your drug every month or every two months, or

whatever it is now. So, you get to remind people on how they are doing. You get a blood test and you make sure, if they are women, that they are not pregnant, and all that kind of stuff.

But there were very few other known hepatotoxins so far where we think you can get around it with monitoring. And I think your question is absolutely right: does the monitoring really happen? Troglitazone called for all kinds of monitoring, but we had lots of fatal injuries anyway. I don't know if that is because nobody did it or if it was because it is too late by the time the transaminase is up; they are going to die.

DR. SENIOR: Well, I practiced 30 years and hardly read labeling except to find out what's the dose. I really never read all the fine print, and I was not alone. That was generally the way it was in those days. I don't know what we can do to get physicians to really pay attention to the labeling. Now the FDA works very hard on negotiating the labeling with the sponsors, and they come to an

agreement, but that doesn't mean anybody is going to read or follow it.

DR. TEMPLE: Those are all good questions. There're other things that go in labeling. What drugs not to use with another drug with is very frequent; how to raise the dose; when to back off. And I think physicians understand some of that. You know, we all know the important stuff goes into a little preliminary summary at the beginning, and I think they read it some, but do they read it enough? I don't know. It is a good question. Mark knows.

DR. AVIGAN: I was going to say that there are studies on monitoring adherence. Actually, in the case of troglitazone there is a litany of studies, some performed by the FDA in the early 2000s about this from the ODS at that time, in collaboration with outside epidemiologists or study people in healthcare systems. It turns out that the system failure, and why it didn't work, was a multi-variable. One was lack of adherence after one or two few months of regular monitoring, as was

asked for in the labeling. The system just lost interest in monitoring the individual. So, the subscription rate for monitoring by patients, by doctors, was very low. And that was shown in a number of studies. After six months, 5-percent adherence rates, because very low yield for the patients who would be bad actors. They have to test a lot of people to find the bad apples.

The second problem with troglitazone, as you pointed out, was the rate of acceleration of injury. When the thunderstorm happens for those individuals, it can be very rapid.

DR. TEMPLE: Right.

DR. AVIGAN: So, you can be happy for a number of months, then get liver toxicity, and there were some documented cases that within a month, which is the interval of monitoring, you were already in a very bad way.

DR. TEMPLE: Right.

DR. AVIGAN: So, the monitoring interval itself was a problem. But that is a characteristic or signature of the drug. So, some drugs could be

valuable if the interval of monitoring was shorter than all the cases where there was an acceleration.

DR. TEMPLE: So, how would we go about knowing? For bosentan, that doesn't seem to be true. You appear to have time to back off.

DR. AVIGAN: Right.

DR. TEMPLE: How would we be able to identify whether it was troglitazone-like or bosentan-like?

DR. AVIGAN: Well, again, we can only speculate, but it would be drug-dependent.

DR. TEMPLE: Right.

DR. AVIGAN: So, you would have to have a natural study of cases, a case study of those individuals who had -- you know, see what the range of accelerations of injury was to see what the interval should be, which is not a good way of doing it because you have to learn from your bad outcomes.

DR. TEMPLE: Right. One would have hoped that some kind of the pre-marketing data would have told you.

DR. AVIGAN: Right, but, unfortunately, the problem is that each case is an anecdote. I wanted to ask you another question.

DR. TEMPLE: Before you do, I want to give one other piece of evidence that sometimes people follow the rules. Terfenadine, as everybody knows, wasn't supposed to cause torsade de pointes when it was used with a 3A4 inhibitor. We put out announcements and told everybody, "Do it before we eventually withdraw the drug." We found and others found that there was about a 90-percent reduction of concomitant use. It wasn't the 100 percent we were hoping for, but it was sharply reduced, which meant some people were paying attention. Now whether that is because of the publicity or our label, I can't tell you the answer to that. But we found that and published it, actually.

DR. AVIGAN: Well, the specter of home monitoring, you know, with a finger-prick test, actually in the far-distant future may solve this problem, where people could just do it at home and not bother. But I wanted to ask you another

question, which is this explosion of information. So, John was asking -- you know, the label is 20 pages long. It could be, actually, 100 pages long. If you really wanted to write a good academic review of a drug and talk about all of the data and all of the different aspects, it could be a much longer distillation than 20 pages. And it’s going to get longer. Let's say demographic characteristics about when you get genetic tested for susceptibility. It could be a very long essay. As we get more and more information about drugs and the nuances of whatever, the question is, as there is an evolving information database, should that all go into the label or should there be a repository of information that informs treatment decision-making for clinicians, because it does inform decision-making in individual cases, that is not part of the label? I have advocated for some time that, actually, it wouldn't be a bad idea to have another site of information which is shepherded or attended by regulatory scientists and academic scientists. It doesn't have to be the

label itself. Because the problem is that the label is always a snapshot; whereas, the information is evolving over time. So, how do we manage the science itself as it expands?

DR. TEMPLE: A formidable question. As you certainly know, we do have a highlight section which is a page that people can get the stuff that is urgent and they are supposed to know. I guess my view is there's all kinds of other -- for example, you won't get a detailed description or analysis of the crucial clinical trials in the label. That is not in there now. You have got to go find the study or something like that or read our reviews. I guess my preference -- and I am not sure everybody would agree -- is that the bulk of the crucial stuff about using the drug should be in the label. We don't list every 3A4 inhibitor. We say you've got to watch out for 3A4 inhibitors, and then, you have got to go to a site to find out what those are, whatever they are. I think that is probably pretty good. It does mean the label is

going to be 12 to 40 pages. But people can choose whether they want to see it or not.

And when you go google things, someone always pulls out the highlights for you in one or another -- whether it is Wikipedia or something else, they have always done that. So, there are short forms, for better or worse, available. Whether there is another place to put it, I mean, if you had a site that specifically relates to the label, so they would know to go there, as opposed to having to search for it, that is not inconceivable. It would be worth thinking about what you put there instead of in the label. I think you can usually write a label that isn't all that extensive. We put junk in them, too. We list adverse reactions that aren't really related and stuff like that.

DR. AVIGAN: My follow-up to that is that the information after the snapshot is done, when you do the label, a week later there is new information. So, the question is, how do you manage to evolve

the information without going through this very formal process of vetting, and so on?

DR. TEMPLE: Oh, very thorny because we don't want people making unsupported claims, do we? Currently, a major First Amendment issue.

DR. SENIOR: Maybe for now we just ask one question. We scheduled the reception for 4:30, but it looks like, that we might finish up a little early. Could we move the reception up a bit?

MS. PAULS: I have wine available at 4:00. (Laughter.)

DR. SENIOR: Okay. Thank you. Well, we don't want to cut the discussion short. The discussion so far has been wonderful. In fact, it is the main thing about the meeting --- not just hearing people lecturing at you, but getting some responses and getting people to argue back and forth. I love it. That is the most important thing, to get people involved in talking about it. So, please carry on.

DR. TEMPLE: So, who looks most argumentative? (Laughter.)

DR. TILLMANN: I have a reputation for that. But, for the monitoring, the electronic medical records are now frequently warning for drug tags, and they probably could include, also, a warning for labeling, where you would actually be able to do something that you cannot renew the prescription for the next month without having the required either ECG or an ANT, if it is for the liver, available.

DR. TEMPLE: So, this would get incorporated into the drugstore requirements. Yes, those are interesting possibilities. If it was something we really thought monitoring was going to be okay, and that was the basis for approving the drug and leaving it on, that seems potentially valuable.

DR. HOWELL: Hi. Brett Howell from the Hamner Institutes. I wondered if I might get your perspective on the scenario where an added benefit is claimed and a monitoring strategy is proposed, but it is in the inpatient setting. So, for example, in the ICU where even daily monitoring is quite easy to do, if that changes the question

of or the benefit/risk equation with respect to monitoring versus these other drugs we are talking about where even monthly is quite onerous for the patient? Thanks.

DR. TEMPLE: Yes, it could. You also have reason to believe the monitoring would actually occur. So, then, the question is, is monitoring likely to be able to prophylax against the problem, which I still think we don't really know very well? I don't know what the difference between bosentan and troglitazone is. Maybe it was the actuality of monitoring, but it also seems like Mark was saying that it has to do with how quickly the bottom drops out.

DR. JONES: Judith Jones. I am not going to put my ex-regulator hat on. I am going to put on a medical community hat. I wanted to thank Mark for making the point about troglitazone. In fact, it was carefully studied. In fact, people did monitor very poorly, and the persistence in monitoring, they read the label and they did not monitor. In fact, they did comply, but not

sufficiently. And that is all very well-documented in the literature.

The second thing: terfenadine. We also did a study of terfenadine and published it in JAMA. In fact, physicians and pharmacists did not necessarily follow the label. In fact, in the study we did in claims data, 50 percent of the pharmacists did not note the drug interaction, despite their warnings, which were available in the pharmacies at the time. And a number of physicians co-prescribed contraindicated drugs. So, there're different types of data on that. The point I really wanted to make was that, in fact -- and you query why the label and all very good information in the label is not heeded. One, a number of medical schools discount the label and the PDR. Two, probably less than 20 percent --

DR. TEMPLE: So, what is their basis for doing anything? What do they recommend instead?

DR. JONES: Pharmacology courses.

DR. TEMPLE: Oh, right.

DR. JONES: But let finish. The point I'm making is, in fact, less than 20 percent of medical schools in this country teach anything that resembles therapeutics. Clinical pharmacists are better equipped to actually look at toxicity. Physicians in most medical schools are still not teaching much in the way of clinical pharmacology and therapeutics. So, you have I don't want to say "a naive," but inexperience audience to review this data and translate it into practice. I think we should really look at ways of doing something about that, because the labels can be excellent. They are widely promulgated, and they are not available. I mean, the readers don't comprehend them and act on them.

DR. TEMPLE: Well, as per the previous discussion, if it is not prescribing two drugs that shouldn't be used together, you really have reason to hope that the pharmacy database can say, "Wait a minute. You're not supposed to do that." Moving on to required monitoring is a big additional step. I am sure there would be complaints from the medical

community and all other kinds of stuff about it. But it certainly is worth thinking about, if the drug was valuable enough so that you think monitoring could save the day.

DR. JONES: Just one final note. Some of the systems in the UK actually do have reminders to monitor in situations like this. As physicians are adopting more and more EMRs, there may be ways of doing that.

DR. TEMPLE: So, when we move to a single-payer system, then we can do that.

(Laughter.)

DR. HICKS: So, Bob, may I just chime-in here? Because she has actually brought up a really excellent point, and we have a lot of people from academia here today. I think the way that we change the culture of not reading the label is to actually incorporate it into the medical curriculum and actually do one class, teach the medical students how to read prescribing information. I had a patient just last week. She was a physician in the hospital and she was

pregnant. She was experiencing a lot of dysrhythmias. I said, "Here, let me show you. There's a great website," you know, Drugs at FDA. We went in and I said, "Here, let me show you how to read this label. First of all, you want to look for any box warnings. Next, you want to go to the highlights." And sure enough, this was an adverse reaction that was reported for this drug product. I really think that, to change the culture -- and I hope there're a lot of people from academia here, that will take this back and talk with colleagues. There should be a class on how to read prescribing information and to use it to your advantage. And only then will people read the label and pay more attention to these kinds of things.

DR. TEMPLE: That class could be paired with one on what is a good clinical trial. (Laughter.)

DR. HICKS: Yes.

DR. TEMPLE: Which is also not taught.

DR. SATINE: B.J. Satine. I have a question. Can we use just label information for DILI to classify the drug for its potential to cause

injury? Or is an overload risk associated with the drug, is that information in the label?

DR. TEMPLE: I'm not sure I understood the question. Are you saying does the label tell us enough about what the experience was, what the findings were, to allow us to classify them properly? Probably not.

DR. SATINE: Based on the DILI information, you can check the label, and based on this information, say some drugs can cause DILI complication or cause some type of DILI? That is the risk?

DR. TEMPLE: Maybe other people do, but I don't know the answer to that. What I realized as I was thinking about this is that I wasn't so sure what we put in each and every case and how consistent it was, how often we said you should get a transaminase, every now and then, or what we do. And I think it is worth a look. But I don't know now.

DR. MADDREY: I have a comment. When do you think we will start cashing-in, if you want to use

that term, on precision medicine, since that is the term this week? All the companies seem to be gathering the appropriate blood for genetic testing. When do you think that the results of genetic testing, which right now are affecting where drugs are used, when do you think the other side of the coin will come into reality and adverse reactions that are determined by genetic testing get equal weight?

DR. TEMPLE: It is a good question. I mean, we know what is going on in oncology.

DR. MADDREY: Yes.

DR. TEMPLE: Okay, that is how we are choosing therapy. I don't know about knowing whether they are going to get a bad outcome, a side effect or not. We are seeing some similar things in cystic fibrosis and stuff like that. But, if you really are trying to say, when are we going to have a choice of cardiovascular medicines, you know, who should get this and who should get that, I don't think we are very far along. In oncology, the sort of lesson is in some sense easy. Cancers are

genetic disorders. So, they have genetic characteristics and you can target therapy toward them. Cardiovascular disease may be, too, but if it is, we don't know it yet. We don't know what set of markers, or whatever, is a good predictor. We know certain characteristics, you know, lipid levels and maybe heart rate. Who knows?

But what you are asking is how are we going to be able to find out who the people who respond to one therapy better than another are genomically or with some other measure. And there are probably a lot of people who are looking. It screams out that in psychiatric disease that ought to be possible. They're genetically-oriented and familially based, you know. Everybody knows that is true for bipolar disease. But we don't have a way of picking out who the responders are going to be. Then, your next question is, well, it was true, did seem to be true -- and maybe Mark knows more -- that for lumiracoxib at least, it looks as if you could identify the people that are going to get in trouble. That is very exciting. I don't know

whether that is going to be true anywhere else, but we found a few cases where that is true.

And, of course, you will never find out unless you look. So, you have to go get a lot of bloods and you have to get the people with toxicity and you have to see if you can find a relationship. And you don't even know what the relationship is going to be. It might be a single gene. It might be a SNP array. I mean, you don't know. Everybody is working their heads off on that, but probably a lot more people than I in the room know what the progress is going to be.

DR. QAZI: I could say a little bit. Sometimes data comes in along with the IND/NDA submissions. And then, it becomes part of the label. There are some examples where there are clear instructions about who should be getting the drug and why certain people shouldn't be getting it, because they are more likely to have adverse reactions. Sometimes they come after the marketing as post-marketing requirement studies. And then, we update the labels with that information. One

example is not in liver toxicity; in abacavir in the HIV scenario, where HLA-B*5701 was associated with adverse events like skin hypersensitivity reaction. Abacavir, hypersensitivity reaction, and that is in the label. And those are followed. But some of them, those that come in along with the IND/NDAs, they oftentimes become companion diagnostics where the drugs and the tests, on the label it says that people who test positive for this biomarker should or should not be getting the prescription. I don't know if it answers your questions. But it has started and I haven't seen it happen in all other areas yet.

DR. TEMPLE: Well, that is the trouble. There are a few of those. The lumiracoxib thing has promise in the same way, but there are very, very few. You know, when we have tried to direct people toward whether to use clopidogrel or not or what dose to use by doing tests that aren't perfect, but are really pretty good, the resentment from the community is extraordinary. They don't want to be bothered by those things. So, it is

work-in-progress and it is very promising. I mean, we have a couple of examples that should inspire us, but it would be hard to say there’re a lot of them.

PARTICIPANT: I'll go first. Thanks. I was interested in the bosentan example because it is a pretty atypical and difficult-to-manage drug, because not only does the bilirubin go up, but frequently it may even go down with continued therapy. So, it is a bit of a special case, and I think that is the context of the monitoring. But what I think that slide raised is the situation where drugs come along; they have this liability. And should they, then, keep that status in the light of newer therapies with no risk? And that is true for the other drugs in pulmonary hypertension, and it is certainly true for labetalol, as far as I can remember. There's sort of pretty funny indications in pheochromocytoma, and certainly the beta dilator and the heart failure indications. There are plenty of other therapies. So, what is the position in that sort of context?

DR. TEMPLE: Well, I think you raise a good question. Do drugs like that, should they persist? Should we rethink them when there are now substitutes? It is a good question. When that came for troglitazone, which actually was, of course, a novel kind of anti-diabetic drug, when two drugs that were just as good came along and it didn't have any advantage anymore, it was gone. Bosentan, at least one of the other drugs is still potentially hepatotoxic, but one maybe not. I think part of the thinking there is you don't want to get rid of the drug that you have maximal experience with. But those are good questions. When has a drug outlived its status is a good question. I am obviously not going to talk too much about that here. But I think we are always thinking about those things. Sometimes a drug should go.

DR. MAYNE: Jim Mayne, Pfizer. I might be able to provide a good followup example on that. Troglitazone, the current example we are discussing, of course, is very dear to our heart at Pfizer. Also, though, sitaxsentan or THELIN,

which was another first cousin of bosentan, I think Bob could provide you with an example of where labeling and monitoring was successful. That drug, while never marketed in the U.S., was marketed in Europe and beyond. It had a label that included close monitoring. It had a patient information program. It had a mandatory registry. So, the ability to gather data in a relatively-complete way was there. In fact, the experience base built that made an argument that it had a less-favorable benefit/risk profile than did the other drugs of the class. So, as sponsor, we made the decision to remove the drug.

The interesting part of the story was, though, that that was easier said than done. Making the decision to remove a drug is not always easy as saying it is too toxic for some people. It was also a drug that was life-saving for some people. It became a very long and difficult exercise to remove it from the marketplace.

So, there are examples where labeling works. There are examples where labeling does not work.

It may be that, where labeling does not work, it is because you are in a large population with a low-frequency event. And as someone else said earlier, people just lose interest in monitoring.

PARTICIPANT: Well, those are points that I was going to make. Now I have a couple of additional ones. Most of the discussion has centered on how to enhance monitoring, read the label more carefully, simplify the label, et cetera, when, in fact, there is precious little evidence, as you have pointed out, Bob, that labeling and monitoring accomplishes anything. Bosentan is often cited as the example. And even there, the case is weak. There are other factors, some of which have been mentioned here. So, I wonder if we should mention that monitoring is a burden. It is a burden on patients. It is a burden especially in the area of co-pays, and it helps drive up medical costs. So, we can't ignore that. I am wondering if we should revisit the issue of how the NTTB drugs are monitored. No chemical monitoring, but symptomatic monitoring. It seems

to work there. And I am thinking that upfront in a label or other information that is disseminated with the summary it says, if there is an issue hepatotoxicity, it says very clearly that, if your patient has symptoms of, quote, "hepatitis," then, just like we do for the NTTB drugs, call your doctor immediately and get tested.

The other thing that we might think about, a sunshine act with regard to monitoring. We have all seen cases where monitoring was required, and then, it turned out not to be a problem. Statins, that poster child; also, the glitazones, the two subsequent glitazones. If we could get monitoring eliminated or at least made less onerous quicker, it might reduce the burden.

DR. TEMPLE: Yes. The question there is what the level of evidence is going to be. It is worth saying we put out something with known toxicity and monitoring to try to minimize it only when we are pretty sure that it does something special. As everybody knows, to get your next clozapine, you have to get a white count, and it

changes over time. But that was because it was considered so valuable, the world needed it, even though there was a risk that some people would die, and some people have died. So, that was worth it. It is always worth figuring out what makes it worth it. And bosentan was easy; there was no other drug. So, you had very little choice. When else would you do that? Just another member of the same class? That doesn't seem very reasonable. So, it is worth thinking about, what makes monitoring, with all its difficulty and cost and questionable effectiveness, a reasonable thing to do? I think what happens is, if a drug looks that valuable, you feel you have to make it available and you are doing your best to minimize the risk, even if you are not sure you are going to be successful.

DR. SENIOR: In the presentation, I thought I heard you say that perhaps bosentan and troglitazone should not have been approved. They were approved.

DR. TEMPLE: No, I didn't say that. I said maybe bromfenac. Bromfenac, not bosentan.

DR. SENIOR: You said bromfenac? But I'm thinking troglitazone should not have been approved.

DR. TEMPLE: There was some evidence of hepatotoxicity before.

DR. SENIOR: As I understand it, the medical reviewer for both drugs was against approval and was overruled. And the reason was stated to be that Congress was putting pressure on the FDA to get drugs approved more quickly, and they did approve a lot of drugs in that year. But eight drugs had to be taken off the market, which is bad for everybody. It is bad for the FDA. It looks like they made a wrong decision. It is certainly a disaster for the company that has an approval and, then, they lose the market from that. And it is even worse for the patient who gets dead.

DR. TEMPLE: Well, I can't vouch for what you said. It might be true, but I don't know it. I actually read a draft version of the review for troglitazone, and I don't think it said, "Don't you dare approve it." And I read the reviews for

bromfenac, and I don't think they said that, either. But they were not as conscious of Hy's Law as I would have been, but that is because I invented it.(Laughter.)

In any event, they did for bromfenac -- all of the cases took a while, and they said this and they said it is only for acute use. You know, you can debate whether that is a plausible restriction for NSAID whose use is mostly for longer terms. But, in any event, we caught it very quickly.

DR. CROSS: Hi. Marcene Cross from Tobira Therapeutics. There are several drugs, as has been discussed here, that have strong labeling relating to hepatotoxicity and the need for monitoring. What strikes me as perhaps different for bosentan is the fact that it is only available under restricted access. And therefore, physicians have to make sure that every four weeks, I believe it is from the labeling, the tests are done and patients can't access their drugs unless even those procedures have been completed. So, I have been wondering whether having that additional layer of

complexity, which I think is just the opposite of what somebody else was saying, that we should make monitoring easier, but by having that additional layer, that that would perhaps help explain why we have seen fewer cases for bosentan.

DR. TEMPLE: Yes, I think it does, but that is a very burdensome distribution system that on the whole people don't like. So, we might use that. If the drug was of great value, we would, and that might help it be better, I am sure you are right. But just putting something like we did for troglitazone and it didn't have a clear requirement, again, either because it doesn't work or because people didn't do it, it didn't accomplish much. Clozapine also, you have to go to your doctor, and, you know, no blood; no drug. So, that is burdensome, but worth it. And the mortality, by the way, I mean, nobody knows really, but agranulocytosis used to be, on the basis of studies from a long time ago, felt to have a mortality in the neighborhood of 10 percent. It has been way, way, way, way less than that. I think

it is because they catch it sooner. But that is a very burdensome system. Not everybody is going to love that.

DR. SENIOR: It looks like Mark Avigan is going to have the last word.

DR. AVIGAN: I was just going to say that, on this issue of closed registries and the following phase, part of what you are doing is managing uncertainty. So, you have a class of drugs and you have a bad actor. You have new drugs in the class you want. You are concerned, but you don't know for sure. In the case of the endothelin receptor antagonist, bosentan was the first. Another one was ambrisentan. It was labeled for hepatic toxicity. And then, at a later point, there was a kind of redress because there was more data in the post-market setting to show that the drug was not really tainted with very much of a signal. So, then, there was a backoff on that label. But, in reality, it was a rationale logic exercise of let's get more data before we back off because we are concerned. So, the question, then, is -- the

absence of information does not imply the absence of risk -- so, the question is, when you have concern and you label upfront in an emerging class, what are the rules for the FDA in terms of when you get the strength of evidence you need to say there's less risk and, then, to back off the label?

DR. TEMPLE: That is a good question. We were familiar with sitaxsentan, which was hepatotoxic also. I think our initial conclusion was to approve, and otherwise, these drugs are likely to be hepatotoxic. And then, for ambrisentan, there was a study in 35 people who had gotten toxic on bosentan, and none of them -- well, maybe one -- most of them didn't have any toxicity on ambrisentan, in contrast to the people who got re-randomized back to bosentan or sitaxsentan who did get toxic. So, really we were sure enough to take it out of the label. Whether that should be done and when is a fair question. I am sure a lot of people think their individualized responses and would be very unhappy with that, all the usual kinds of concerns.

DR. SENIOR: Well, it is now four o'clock, and I think we are going to thank Bob Temple for a very thoughtful presentation and for leading such a stimulating discussion. Thanks, Bob. (Applause.) To me, the best part of these meetings is the open discussion that takes place after the presentations, not just the canned presentations, but the open back-and-forth between the people who have registered and come. They have come a long way to be here and to have a chance to hear, but, also, a chance to speak their own minds. So, with that, we will take a break, have a reception, have some dinner, and a little rest. And don't forget to come back; save your energy for one more shot with Paul Watkins at eight o'clock. 4:03 pm