Liver Test Elevations in Patients on Placebo

Robert W. Tipping
Director, Clinical Biostatistics
MERCK Research Laboratories

Heptox Special Interest Group
Silver Spring, Maryland
March 26-27, 2008

Slide 1

MR. TIPPING: All right. Good morning, everybody and thank you, Paul and John, for organizing this and having me here. When John called several months ago and asked me to do this talk, I thought, well, this is work that John and I did several years ago using some Merck data. Actually I presented it to this group a while back, and at that time the group was about 50 or 60 people.

DR. SENIOR: Maybe about 30.

MR. TIPPING: Yeah, maybe 30. So when John called and said, maybe we could dust those slides off, I thought, well, you what's the interest going to be, the same 30 or 40 or 50 people might be there.

So it is great to see that through the years this meeting has grown in size. Interest has really gone up. The facilities keep getting better every year, thanks to John and to Lana. So it is nice to be back and refresh people. So I apologize for those of you that maybe have seen some of this information a few years back, but I think it is interesting and follows nicely along from some of the information Naga was just speaking about.

I'm going to talk a little bit about liver test elevations in patients on placebo. My name is Bob Tipping. My badge says Robert. Please don't call me that. Call me Bob. I'm a director in a clinical biostatitics unit at Merck. So probably I, unlike a lot of the other speakers here, am not a medical doctor. Statistics is my specialty and my interest in this is really measurement characteristics of some of these biomarkers.

Acknowledgements

PhRMA / FDA Hepatotoxicity Steering Committee

Dr. John Senior, FDA

Dr. Peter Honig, MRL

Posthumously, Dr. Harry Guess

Slide 2

I do want to acknowledge some people who have motivated and helped with this work. Certainly this group which was formally known as the PhRMA-FDA Hepatotoxicity Steering Committee and now is I guess is the Special Interest Group. Dr. John Senior, sitting up here from the FDA, was a great help. In fact, a few years back when we were doing this, John and I spent many months, about every other Friday afternoon together. For the first month or so, he educated me on some of the clinical issues, and then we really dove into our data set.

Dr. Peter Honig, formerly of the FDA, now at Merck, was very helpful in guiding some of this work and in clearing some of my other priorities out of the way so that I could do some of the things that we were able to do.

And then finally, I want to acknowledge Dr. Harry Guess, who unfortunately passed away a few years back, but was a great motivator of this and very interested in this topic.

Screening for DILI
Current Practice

- Typical statement from drug label
  - It is recommended that liver function tests be performed at baseline and periodically thereafter (e.g., semiannually) for the first year of treatment …
  - If the transaminase levels show evidence of progression, particularly if they rise to three times the ULN and are persistent, the drug should be discontinued.

Slide 3

So what you see here is a typical statement from drug labels across many classes where it says, it is recommended that liver function tests be performed at baseline and periodically thereafter, for instance, semiannually for the first year of treatment. And if the transaminase levels show evidence of progression, particularly if they rise to three times the upper limit of normal and are persistent, the drug should be discontinued.

But what de we know about the measurement characteristics of these transaminase levels? And as Naga got into it in greater depth, what’re the issues of ALT and AST and the fact that they are found in many other organ systems of the body: muscle, heart, kidney? And then what's the value of the transaminases, which are really measures of liver injury as opposed to other parameters which are more a measure of liver function such as bilirubin?

Screening for DILI
Problems with Current Practice

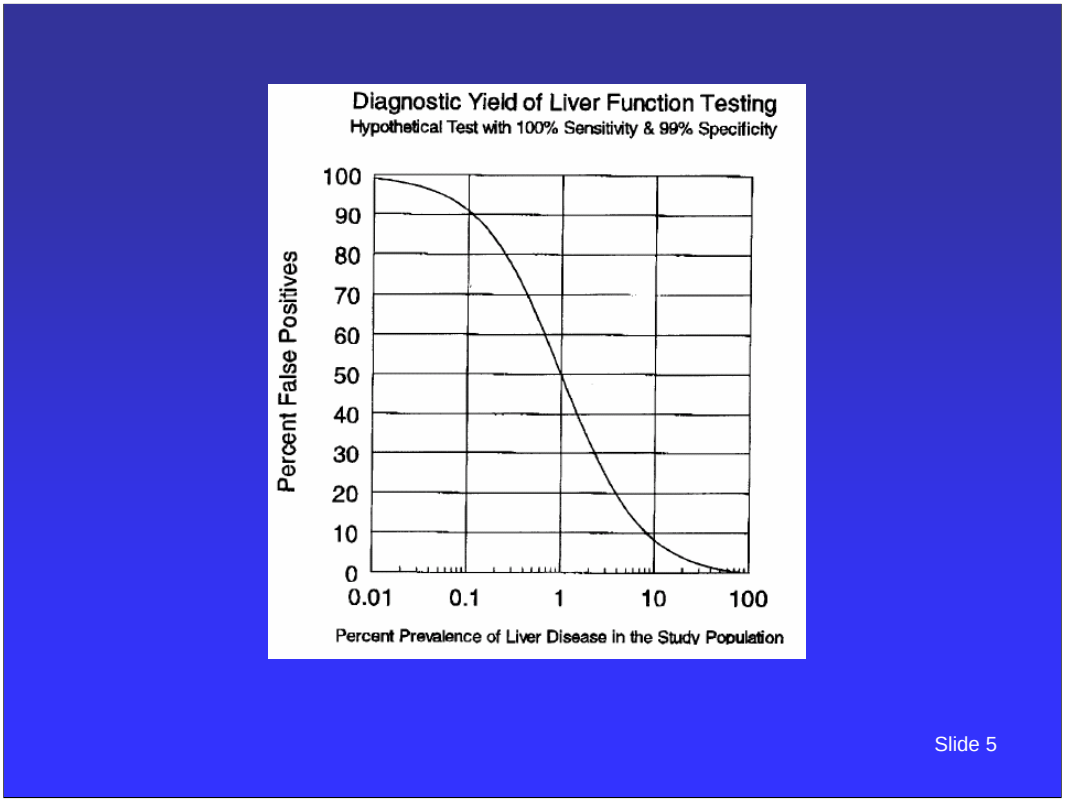
- Abnormal liver chemistries can arise from many causes
  - acute and chronic viral infections
  - alcohol consumption
  - medication use
  - comorbid disease
  - strenuous exercise
- Idiosyncratic hepatotoxicity typically a rare event
- Elevated transaminase levels have questionable sensitivity, specificity and PPV for predicting DILI

Slide 4

So to just talk a little bit about what current practice and current drug labels are, what are some of the problems with them? Well, again as pointed out in previous talk, abnormal liver chemistries can arise from many causes, some of which I've listed here. Viral infections, alcohol consumption, medication use, and this is not just limited to prescription drug use but in the world we live in today, with the availability of many OTC therapies and nutraceuticals, these things can also impact transaminase levels. Comorbid disease and strenuous exercise, are these transaminase levels increasing because of something going on in the muscle or elsewhere, possibly? All of these have impact on specificity of transaminase levels for true drug-induced liver injury.

The other problem we deal with is that idiosyncratic hepatotoxicity is a very rare event. This low incidence has a great impact on positive predictive values. And then finally, elevated transaminase levels do have questionable sensitivity, specificity, and positive predictive value for predicting drug-induced liver injury.

We chose to investigate the issue of how predictive elevated serum transaminase levels were for non-drug-related detection of sporadic liver diseases, in a population of almost 3300 patients followed with serial measurements periodically for five years while taking only placebo. We wanted to know if Hy Zimmerman's observation about jaundice adding information of value in drug-induced hepatocellular injury might also apply to detection of serious liver disease.



This graph is a little bit hard to read, but it does begin to give you a sense of the impact of the measurement characteristics of transaminases on the issue that we're actually trying to get at, detection of serious liver injuries. So to orient you a little bit with this, this graph shows the diagnostic yield of liver testing for a hypothetical example where the test has 100 percent sensitivity and 99 percent specificity for predicting the amount of interest. These would be very, very good measurement characteristics, and I don't think anybody in this room thinks that what we can do now comes anywhere close to this.

What you see on the X axis is the percent prevalence of liver disease in the study population: all the way to the right side if there was 100 percent prevalence, versus all the way to the left if it was 1 in 100 or 1 in 1,000 or 1 in 10,000. On the Y axis is the percent of false positives which is one minus the specificity. The one take off from this message is that we're really operating on the left side of this graphic where if we say that the incidence of liver disease in the study population is 1 in 10,000, and is probably lower than that, then 98 percent of our positive tests are going to be false positives.

So think about some of the drug labels and some of the drug classes that have this type of labeling and look at transaminases for predicting drug-induced liver injury. For instance, the statins that are prescribed for millions of people, you begin to understand the economic impact of this. Potentially there is also negative impact on mortality and morbidity from cardiovascular disease when you consider the numbers of people that are discontinued from their therapy because of false positive elevations in transaminase levels.

PhRMA/FDA/AASLD
Drug-Induced Hepatotoxicity White Paper
Postmarketing Considerations

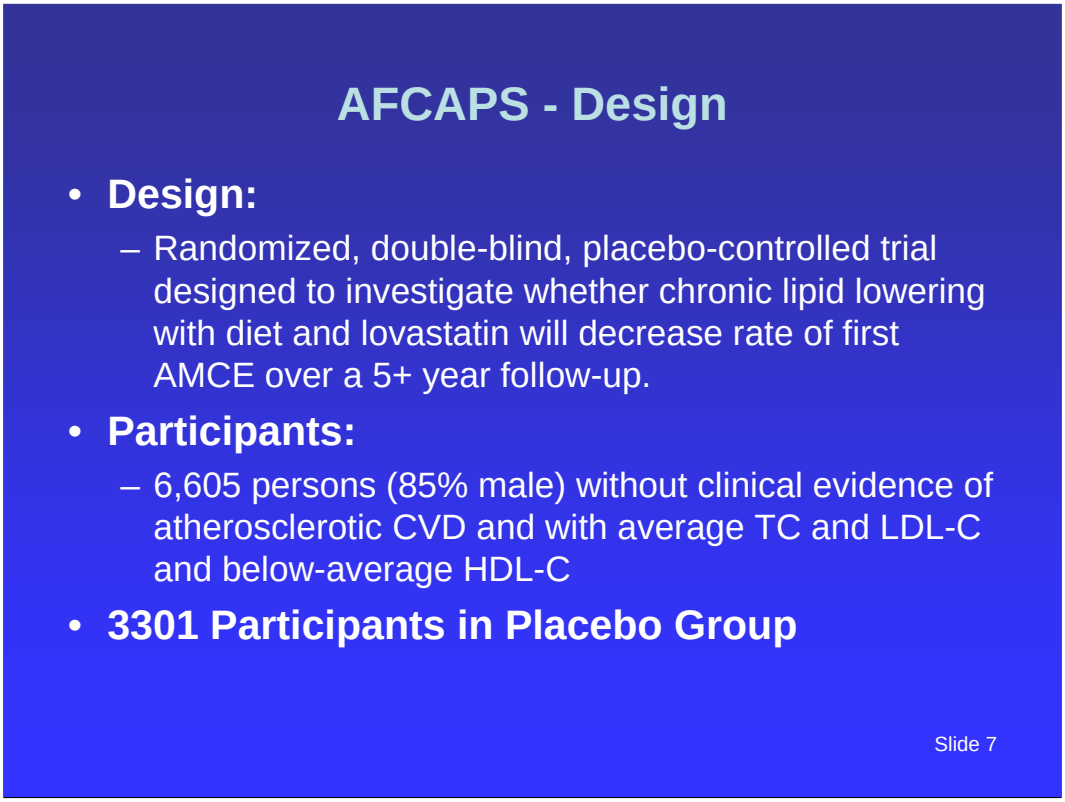
Initiative 4

- Using the placebo arms of long-term controlled trials in various disease-specific populations:
  - Quantify the frequency of transaminase abnormalities and
  - Assess the intrasubject and intersubject variability in transaminase levels

Slide 6

So coming back now to the motivation for this, this work actually came out of the initial meeting of this group which took place I believe in 2001 in Chantilly, Virginia, where these concerns were addressed. One of the things that was discussed at that meeting was getting a better idea of the measurement characteristics of some of these biomarkers and indicators for drug-induced versus disease-induced liver injury.

From that meeting and the preceding White Paper of November 2000 on Postmarketing Considerations, Initiative 4 (see www.fda.gov/cder/livertox/postmarket.pdf). I believe there were 8 that came out of that meeting but Initiative Number 4 was using the placebo arms of long-term controlled trials in various disease-specific populations to quantify the frequency of transaminase abnormalities. It was proposed to assess some of the measurement characteristics, namely the intra-subject subject and inter-subject variability of these levels. And it was at this point that John and Harry Guess got me involved to begin to look at some of the Merck trials and some of the large placebo databases that we had, because they seemed to be good datasets.



The trial that we selected to look at is called AFCAPS, which stands for the Air Force Coronary Atherosclerosis Prevention Study. In the next few slides, I'll just go briefly through the design of this trial. The AFCAPS was designed as a randomized, double-blind placebo-controlled trial to determine whether chronic lipid lowering with diet and with lovastatin, otherwise known as Mevacor, would decrease the rate of first acute major coronary events over a five-plus-year follow-up period. AFCAPS enrolled over 6600 patients, predominantly male, 85 % male, without clinical evidence of athersclerotic cardiovascular disease, with average total cholesterol and LDL cholesterol values and below-average HDL. It was a population that was generally healthy. About half of those participants were in the placebo group, slightly over 3300 patients. So that was really a very special trial in that it gave us a large placebo group followed over a long period of time.

I'm just going to pause here to say, you know, what you're going to see over the next several slides points to the value of large placebo databases like this. There's sort of an ongoing PhRMA initiative right for companies coming together and actually pooling studies like this to get even larger groups followed for lengths of time, and that initiative is actually kind of competing with some other PhRMA initiatives. So for those of you who go back and have the ability to impact your company's thinking on these issues, hopefully some of the things that I'm going to tell you about AFCAPS will strike a chord with you. You can consider and expand on it and ask: what if we could combine 10 such trials such as AFCAPS? how much more could we understand about some of these things? I'll stop with the commercial now.

AFCAPS - Design

- **Lab Evaluations:**
  - Serum chemistry (ALT, AST, TBL, ALP, & CPK) evaluated at:

Baseline	(3 exams):	Wks –4 & –2, Day 1
Year 1	(8 exams):	Wks 6, 12, 18, …, 48
Years 2-5	(9 exams):	Wk 60 & Mos. 18, 24, …, 60

- **Relevant Inclusion Criteria:**
  - Men 45 to 73 years; Women 55 to 73 years
  - ALT/AST not >20% above ULN at Wks –4 or -2
- AJC (1997), JAMA (1998), AJC (2001)

Slide 8

Continuing on with the AFCAPS design, one of the other real values of this trial were the number of laboratory evaluations that were being performed. So serum chemisties, including those parameters that you see there, were evaluated about 20 times in each subject: 20 scheduled times over the course of the 5-year trial along with additional unscheduled measurements; 3 exams occurring at baseline, 8 exams occurring within the first year at about 6-week intervals, then 9 more exams occurring throughout years 2 through 5.

Relevant inclusion criteria for AFCAPS included enrolling men 45 to 73 years of age, women 55 to 73 years of age. They were only included if their ALT and AST at baseline or at screening I should say, week minus 4 and week minus 2 evaluations, were not to be more than 20 percent above the upper limit of normal. For those of you who are interested in learning a little bit more about AFCAPS, I've just listed three of the publications that came out. First, a 1997 publication that was a design paper, then a JAMA publication where primary efficacy results were discussed, and third a 2001 publication in which additional long-term safety findings from AFCAPS.

So from AFCAPS, what do we know about quantifying the frequency of transaminase abnormalities? The next couple of slides are fairly data-intense. So I'll try to go through them slowly.

The first thing I want to point out is that AFCAPS represents a very rich placebo database in which you can see here -- what this slide shows is ALT, AST and TBL which stands for total bilirubin, and if you look within each of those three variables across the baseline, year 1 and year 2 through 5 rows, you see on the order of about 58-60,000 tests being performed in these 3300 subjects. So it was a very rich database to look at these sorts of measures.

The next thing I want to point out, again recalling the inclusion criteria for AFCAPS, is that people basically had to be normal with regard to their -- within normal ranges for ALT and AST, yet you see at the baseline value which occurred on day 1, you're still seeing 263 individuals, for example, with ALTs between 1 and 2 times the upper limit of normal, 13 at 2 to 3 times upper limit of normal, and 2 that were 3 to 5 times the upper limit of normal. Thay hadn't been exposed to any drug, hadn't been exposed even to placebo. This was just from week -4 to week -2 when they were no more than 20 percent above normal to 2 to 4 weeks later at the Day-1 exam elevation. Why did this happen? It was certainly not a drug related event.

The other thing I want to point out from this is: looking down through baseline, then the year 1 line and then years 2 through 5, it's very much a pattern of the more you look, the more you're going to see. And then finally, I just know every time John and I would get together and talk about bilirubin, he would give me the little lecture on Gilbert's syndrome, which is actually something that he and I have ourselves, occurring at about 5 to 7 percent prevalence in the population.

AFCAPS Placebo Group
Serum Chemistry Data

Parm	Time	# Pats	# Tests	Distribution of Participants by Max Elevation (xULN)					
				<=1x	>1-2x	>2 3x	>3-5x	>5 10x	>10x
ALT	BL	3301	10118	3023	263	13	2	0	0
	Yr 1	3248	24640	2700	504	32	5	5	2
	Yrs 2 5	2861	23322	2325	447	63	17	7	2
AST	BL	3301	10120	3204	92	1	3	1	0
	Yr 1	3248	24650	3013	210	15	4	5	1
	Yrs 2 5	2861	23316	2530	294	21	9	6	1
TBL	BL	3301	10120	2832	444	24	0	0	1
	Yr 1	3248	24629	2506	696	45	1	0	0
	Yrs 2 5	2861	23067	2215	607	33	3	3	0

TBL = Total Bilirubin

Slide 10

The first thing I want to point out is that AFCAPS represents a very rich placebo database in which you can see here -- what this slide shows is ALT, AST and TBL which stands for total bilirubin, and if you look within each of those three variables across the baseline, year 1 and year 2 through 5 rows, you see on the order of about 58-60,000 tests being performed in these 3300 subjects. So it was a very rich database to look at these sorts of measures.

The next thing I want to point out, again recalling the inclusion criteria for AFCAPS, is that people basically had to be normal with regard to their -- within normal ranges for ALT and AST, yet you see at the baseline value which occurred on day 1, you're still seeing 263 individuals, for example, with ALTs between 1 and 2 times the upper limit of normal, 13 at 2 to 3 times upper limit of normal, and 2 that were 3 to 5 times the upper limit of normal. Thay hadn't been exposed to any drug, hadn't been exposed even to placebo. This was just from week -4 to week -2 when they were no more than 20 percent above normal to 2 to 4 weeks later at the Day-1 exam elevation. Why did this happen? It was certainly not a drug related event.

The other thing I want to point out from this is: looking down through baseline, then the year 1 line and then years 2 through 5, it's very much a pattern of the more you look, the more you're going to see. And then finally, I just know every time John and I would get together and talk about bilirubin, he would give me the little lecture on Gilbert's syndrome, which is actually something that he and I have ourselves, occurring at about 5 to 7 percent prevalence in the population.

AFCAPS Placebo Group
Serum Chemistry Data

Parm	Time	# Pats	# Tests	Distribution of Participants by Max Elevation (xULN)					
				<=1x	>1-2x	>2 3x	>3-5x	>5 10x	>10x
ALP	BL	3301	10111	3256	44	0	1	0	0
	Yr 1	3248	24597	3178	70	0	0	0	0
	Yrs 2 5	2861	23003	2792	63	3	0	3	0
CPK	BL	3301	10123	2605	566	92	29	7	2
	Yr 1	3248	24703	2114	864	149	73	36	12
	Yrs 2 5	2860	23016	1892	726	145	66	22	9

Slide 11

This is a very similar data table, now looking at alkaline phosphatase (ALP) and creatine phosphokinase (CPK). We do see fewer elevations with ALP than we saw with some of the other parameters but we still do see the phenomenon of that the more you look, the more you're going to see. For CPK we see numerous elevations during the course of the trial, which gets back to Naga's point about what's actually causing this, maybe exercise-induced. It's more indicative of muscle injury than anything going on with the liver, but it's important I think when you're trying to understand why an individual may be having elevations in transaminases, you need to understand what's going on with some of these other variabless, too. You can't just say, well, there's an ALT elevation, so therefore something must be going on in the liver.

AFCAPS Placebo Group
Serum Chemistry Summary

Cumulative Incidence of Maximum Elevations of Blood Chemistry Parameters Relative to ULN
(Post-Randomization)

Parameter	Normal Range	Defn of Elevation	n	(%)
ALT (IU/L)	0 40	>3X ULN	38	(1.17)
AST (IU/L)	0-37	>3X ULN	26	(0.80)
TBL (mg/dl)	0.0-1.0	>2X ULN	65	(2.00)
ALP (IU/L)	M 39-48 yrs 58-137 M 49-73 yrs 59-146 F 49-68 yrs 56-148 F 69-73 yrs 64-162	>2X ULN	6	(0.18)
CPK (IU/L)	M 24-195 F 24-170	>5X ULN	77	(2.37)

Slide 12

Now let's start to summarize these data and drill down a little bit more. We're looking at serum chemistries but summarizing them to some degree. What we're looking at here is the cumulative incidence of the maximum elevations of blood chemistry variables relative to the upper limit, but normal during the post-randomization period. Okay.

So what we see for ALT, normal range of 0 to 40, and the definition of an elevation that John and I used was 3 times the upper limit of normal. We saw 38, or a little over 1%, of the population of placebo patients showing this level of ALT elevation over a 5-year period. For AST, normal range of 0 to 37, a slightly lower incidence, 26 patients, slightly less than 1%. For total bilirubin, normal range 0 to 1, elevation of 2 times the upper limit of normal in 65 patients with that, just about 2%. For ALP, by gender and by age, not manyof elevations here, only 6 out of 3300 patients, and then finally CPK, the parameter that showed the most frequent elevation, 77 or about 2.5%. So we see lots of elevations here. Are they actually signals of trouble or of disease, or are they just normal variations?

AFCAPS Placebo Group
Serum Chemistry Summary

Incidence of Concurrent Elevations of Two Blood Chemistry Parameters Relative to ULN
(Post-Randomization)

	ALT >3X ULN n (%)	AST >3X ULN n (%)	TBL >2X ULN n (%)	CPK >5X ULN n (%)
ALP >2X ULN	3 (0.09)	3 (0.09)	3 (0.09)	0
ALT >3X ULN		20 (0.62)	5 (0.15)	1 (0.03)
AST >3X ULN			5 (0.15)	3 (0.09)
TBL >2X ULN				0

Slide 13

So now let's drill down to the next level and look at, in our serum chemistry summaries, at the incidences of concurrent elevations of two blood chemistry parameters relative to the upper limit of normal. So looking at these five different variables and looking at concurrent elevations of any two of them, what we begin to see here is evidence of increased specificity. There're fewer numbers here, smaller numbers in people showing these elevations. We're down now to people that had concurrent elevations of ALT and AST down to 20. You recall it was 38 had ALT elevations, 26 had AST elevations but only 20 had concurrent elevations of both of those.

Possibly more importantly, looking at concurrent bilirubin elevations, five of the individuals had concurrent elevations of ALT and bilirubin and the same five incidentally also had AST and bilirubin. So those five individuals had all three of those parameters elevated above upper limit of normal.

Finally, looking at the ALT and the AST rows, the last column, at concurrent elevations of those two parameters with CPK, very, very small numbers of people, one and three respectively, and is this really pointing to something that maybe is more muscle-related rather than liver-related?

"In his 1978 book, "Hepatotoxicity," Dr. Hyman Zimmerman noted that the combination of pure hepatocellular injury (transaminase elevation without much ALP elevation) and jaundice was particularly ominous, with about 10-15% of such patients who showed such findings as a result of drug-induced injury going on to die. Recent experience has borne out the observation that the combination of transaminase and bilirubin elevation often predicts the occurrence of severe injury in some patients. The idea that the combination of elevated transaminase(s) and total bilirubin has ominous implications has come to be dubbed "Hy's Law" at the Agency". [Clinical White Paper]

Working definition for Hy's Rule: >3x ULN transaminase elevations concurrent with >2x ULN total bilirubin

Does Hy's Rule improve specificity for identifying drug-induced liver injury?

Slide 14

So now we get around to Hy's Rule or Hy's Law and I'm sure John will have something to say about which that should be and Bob Temple's here somewhere in the room. We'll have an interesting discussion between Bob and John about whether it's a rule or it's a law and how it came about but quoting from the 1978 textbook by Dr. Hyman Zimmerman, he noted that "The combination of pure hepatocellular injury and jaundice was particularly ominous, with about 10-50% of such patients who showed such findings as a result of drug-induced injury going on to die." He meant transaminase elevations without much alkaline phosphatase elevation. Recent experience has borne out the observation that the combination of transaminase and bilirubin elevations often predicts the occurrence of severe injury, or disease, in some patients. The idea that the combination of elevated transaminase(s) and bilirubin has ominous implications really was Hy Zimmerman's observation.

So part of what John and I did a few years back was to try to use the AFCAPS data to begin to look at the combination of ALT and TBL elevations and the predictive value of it for liver disease, in patients on placebo and therefore not drug-induced. Our working definition as we applied it to the AFCAPS data was an elevation three times upper limit of normal for transaminase together with an elevation two times the upper limit of normal for total bilirubin. So the question was, does the combination of elevated TBL with elevated ALT improve the specificity for identifying disease-induced liver injury?

AFCAPS Placebo Group
Serum Chemistry Summary

- Reported results
  - 44 participants (1.35%) w/ elevations >3x ULN ALT/AST
  - 11 participants (0.34%) with confirmed elevations >3x ULN
  - 5 of 11 experienced elevation during 1st year of treatment
- ALT/AST and TBL results
  - 5 participants (0.15%) with >3x ULN ALT/AST elevations along with >2x ULN TBL
  - 1 additional participant with >2x ULN ALT/AST elevation along with >2x ULN TBL

Slide 15

The serum chemistry summary from AFCAPS, the reported results, if you go to some of the publications that I just talked about and look at what the reported prespecified results were for this, 44 patients or about 1.4 percent had elevations 3 times the upper limit of normal in either ALT or AST, 11 of those 44 had those elevations confirmed which was something that was specified in the protocol, 5 of those 11 had that confirmed elevation within the first year of treatment. So about half of them occurred in the first year. The other half occurred years 2 through 5, which is interestingly enough very consistent with the frequency of the measurements. Half the tests that we did were in the first year. So the more you look, the more you see.

Now moving down to the bottom part of this slide, not prespecified, this is the work that John and I did, so what happens if you look at this database and look now for so-called ALT-TBL combination patients, as least as we've defined them. What we saw were 5 participants who actually met this criterion, and all 5 had serious liver disease requiring hospitalization and 1 died of infiltrating amyloidosis. There was one more patient, a sixth, who showed elevated AST and ALT followed by TBL elevation after being taken off study, who later died of metastatic cancer, as you'll see in future slides.

AFCAPS Placebo Group
Serum Chemistry Summary

Case #	Gender	Age	Confirmed ALT/AST 3x ULN Std Day	ALT/AST and TBL Std Day	Clinical Finding
1	M	61		919 †	amyloidosis, liver failure
2	M	51	1090		
3	M	56	2058		
4	M	70		715 ‡	cholecystitis
5	M	49		1723	cholecystitis
6	M	62	567		
7	M	72	127	317	hepatitis, chronic
8	M	59		1338 ‡	metastatic colon cancer
9	M	57	1695	1702	hepatitis A
10	M	46	69		
11	M	55	396		
12	F	65	86		
13	F	59	720		
14	F	68	43		
15	F	64	211		

† ALT and AST 2.4x ULN; concurrent ALP 6.6x ULN

‡ Concurrent ALP 2.4x ULN (Case #4) and 6.7x ULN (Case #8)

Slide 16

In this slide, I present 15 of the 3300 patients from AFCAPS and I hope people can see this well. What it represents is people who either had confirmed elevations of ALT or AST, and then the 6 that met the criterion pof combined ALT and TBL elevations. In the entire database the 3248 reviewed in an extensive review performed by John, these are actual clinical findings of liver outcomes, not drug-induced liver injury because these patients were taking placebo but disease induced serious liver diseases that revealed with great specificity by the biomarkers ALT and TBL.

Review of the data included our case report form information and WAES (World Adverse Event Study) reports, and then John went in and looked at comments written in the margins and really did an extensive review to diagnose these cases. I do want to point out that the double dagger indicates two additional cases where there was increase in ALP.

And then returning to the 11 cases in the 4th column, so this was what the protocol prespecified as our criterion (confirmation of serum transaminase elevations by prompt repeat testing) for looking at potential liver injuries. What you see is that of the 11, only 2 of them actually identified cases in whom there was a true clinical finding. The other nine were false positives, depending on serum transaminases alone.

The column for combined ALT and TBL elevations really identified no false positives. Every time that someone was flagged, there actually was some relevant clinical finding. So in this particular case, there was a positive predictive value of 100 percent.

Definition of Metrics

- Sensitivity
  - In those with the condition, what is the probability of a positive test
- Specificity
  - In those without the condition, what is the probability of a negative test
- Positive Predictive Value (PPV)
  - In those with a positive test, what is the probability of having the condition

Real quickly, the definition of metrics that we're going to look at, sensitivity and those people with the condition, what is the predictive power of a positive test? For those without any serious liver injury, what was the predictive power of a negative test?

AFCAPS Placebo Group

Sensitivity, Specificity, and PPV of Various Criteria for Detecting True Cases (n=6) N=3248

Test Criteria	# of "+" Tests	# Cases Detected	Sensitivity	Specificity	PPV
(ALT or AST)>3xULN	44	5	83.3%	98.8%	11.4%
ALT>3xULN	38	5	83.3%	99.0%	13.2%
(ALT or AST)>3xULN, confirmed	11	2	33.3%	99.7%	18.2%
(ALT or AST)>5xULN	19	5	83.3%	99.6%	26.3%
(ALT or AST)>5xULN, confirmed	4	2	33.3%	99.9%	50.0%
(ALT or AST)>3xULN & TBL>2xULN	5	5	83.3%	100%	100%

Slide 18

What this slide shows is various combinations of these biomarkers and the number of people that actually met those conditions and the number of the cases that were actually detected. I'm running a little bit short on time here, so I'll move through this pretty quickly. The first row shows ALT and AST, ALT or AST elevated 3 times the upper limit of normal but without confirmation, 44 tests, 44 positive tests detected 5 of our cases. So reasonable sensitivity and specificity and a positive predictive value of 11 percent.

Coming down to the third line now, I'll try to move through this quickly, this was our prespecified finding: ALT or AST elevated but it had to be confirmed. So what we've done here is we've eliminated a lot of the false positives. It comes from 44 down to 11, the total number of cases but we've also now only detected 2 of the 6 cases for an unsatisfactory sensitivity of only 33 percent.

And then finally jumping all the way to the last row, which is our combination ALT-TBL definition, both elevation of ALT or AST along with bilirubin, 5 positive tests identifying 5 of the cases, 5 of 6 cases, for an 83 percent sensitivity, 100 percent specificity and 100 positive predictive value. The sixth case met the criterion, but after has was taken off study because of transaminase elevations.

I do want to point out that while this is a large placebo database, 3300 patients, we still are only talking about 6 cases, and incidence of detected serious liver disease of only 1 per 550 (0.18%) in 5 years. So a small number of cases actually has a great impact, which comes back to what if we could combine AFCAPS with 9 or 10 other trials like this?

Assess the intrasubject and intersubject variability in transaminase levels

Slide 19

So I'll just take 30 more seconds, for those of you who are interested, I do have some information on measurement characteristics, inter- and intra-subject variability, coefficient of variation, and interclass correlations.

Definition of Metrics

- Coefficient of Variation (CV)
  - defined as intrasubject standard deviation expressed as a percent of the mean
  - important metric for the stability of a single measurement
- Reliability Coefficient (R)
  - quantified by the intraclass correlation coefficient (ICC)
  - defined as the proportion of true (intersubject) variability relative to the total variance
  - equal to the correlation of measurements within same subject
  - important metric for the stability of measurements over time when no true change has occurred
- Outliers can dramatically affect both metrics

Slide 20

So I'll just take 30 more seconds, for those of you who are interested, I do have some information on measurement characteristics, inter- and intra-subject variability, coefficient of variation, and interclass correlations.

(quickly shown, without comment)

Data Exclusions

- Visit data from 9 participants excluded due to clinical findings that could affect a change in serum chemistry measurements
- Twelve additional data points (7 ALK, 1 TBL, 4 CPK) identified as outliers and excluded.
  - Lab error?
  - Physiologic change associated with unreported clinical event?

So I'll just take 30 more seconds, for those of you who are interested, I do have some information on measurement characteristics, inter- and intra-subject variability, coefficient of variation, and interclass correlations.

(quickly shown, without comment)

AFCAPS Placebo Group

Measurement Characteristics of Serum Chemistry Parameters

Year 1

Parm	Gender	mean	intra subject variance	inter subject variance	C V	I C C
ALT	Total	20.5	34.7	55.5	28.7	0.62
	Male	23.1	36.8	59.0	26.3	0.62
	Female	17.9	23.3	36.0	27.0	0.61
AST	Total	20.2	15.0	19.0	19.2	0.56
	Male	21.0	15.5	19.1	18.8	0.55
	Female	19.4	12.1	18.7	18.0	0.61
TBL	Total	0.60	0.033	0.053	30.3	0.61
	Male	0.68	0.035	0.058	27.5	0.62
	Female	0.52	0.022	0.024	28.9	0.51

So I'll just take 30 more seconds, for those of you who are interested, I do have some information on measurement characteristics, inter- and intra-subject variability, coefficient of variation, and interclass correlations.

(quickly shown, without comment)

AFCAPS Placebo Group

Measurement Characteristics of Serum Chemistry Parameters

Year 1

Parm	Gender	mean	intra subject variance	inter subject variance	C V	I C C
ALP	Total	82.1	52.6	379.4	8.8	0.88
	Male	79.0	50.9	363.9	9.0	0.88
	Female	85.2	62.1	468.1	9.3	0.88
CPK	Total	104.4	2558	3621	48.4	0.59
	Male	125.2	2808	4056	42.3	0.59
	Female	83.6	1166	1148	40.8	0.50

So I'll just take 30 more seconds, for those of you who are interested, I do have some information on measurement characteristics, inter- and intra-subject variability, coefficient of variation, and interclass correlations.

(quickly shown, without comment)

AFCAPS Placebo Group

Measurement Characteristics of Serum Chemistry Parameters

Years 2-5

Parm	Gender	mean	intra subject variance	inter subject variance	C V	I C C
ALT	Total	20.9	47.2	59.7	32.8	0.56
	Male	23.4	49.0	63.3	29.9	0.56
	Female	18.5	36.9	38.1	32.9	0.51
AST	Total	21.4	19.8	20.7	20.8	0.51
	Male	22.2	19.8	19.9	20.1	0.50
	Female	20.7	19.8	26.1	21.5	0.57
TBL	Total	0.63	0.023	0.053	24.3	0.70
	Male	0.71	0.025	0.057	22.2	0.70
	Female	0.54	0.012	0.028	20.5	0.69

So I'll just take 30 more seconds, for those of you who are interested, I do have some information on measurement characteristics, inter- and intra-subject variability, coefficient of variation, and interclass correlations.

(quickly shown, without comment)

AFCAPS Placebo Group

Measurement Characteristics of Serum Chemistry Parameters

Years 2-5

Parm	Gender	mean	intra subject variance	inter subject variance	C V	I C C
ALP	Total	80.4	62.6	348.7	9.8	0.85
	Male	78.9	57.0	339.1	9.6	0.86
	Female	81.9	94.7	404.0	11.9	0.81
CPK	Total	100.4	2537	3510	50.1	0.58
	Male	123.0	2845	3945	43.4	0.58
	Female	77.7	778	1036	35.9	0.57

So I'll just take 30 more seconds, for those of you who are interested, I do have some information on measurement characteristics, inter- and intra-subject variability, coefficient of variation, and interclass correlations.

(quickly shown, without comment)

Conclusions

- Hy's Rule (>3x ULN ALT/AST concurrent with >2x ULN total bilirubin) appears to be highly specific for liver disease
- Reliability of ALT, AST, TBL, and CPK is modest
- Improvements to reliability can be achieved by
  - repeat measurements
  - adding a second indicator (as in Hy's Rule)

Slide 26

I'll jump right to my conclusion slide. Real quickly for 30 more seconds, one thing that I concluded from this is that serum transaminase elevations are not a disease, although it's not up here. The AST did not add much as a predictor over ALT alone, but was redundant. The combination of ALT and TBL elevations, as is defined here, was highly specific for liver disease, although I didn't show you much of the data on that. On the second bullet point there, the reliability of ALT, AST and TBL, and CPK is modest at best. The ALP is much better but again reproducibility and reliability of some of these tests are not all that good.

We can obtain improvements in the reliability of the tests by doing repeat measurements, as we did in the AFCAPS trial. We found that adding a second indicator as in the case with ALT-TBL combination can really improve the reliability and unpredictability of testing.

So I'll just close, again pointing to the AFCAPS placebo database of 3300 patients, by reconizing although it's not up here, the tremendous value that can be obtained by exploration of placebo databases.

Thank you.

(Applause.)