Alanine Aminotransferase: A Clinical and Regulatory Tool for Detecting Liver Injury–Past, Present, and Future

JR Senior[1]

Assay of the serum activity of the enzyme alanine aminotransferase (ALT) has become the primary screening tool for detecting acute liver injury. But what does an elevated value mean? Not what it is too often mistakenly believed to indicate. It is not a test of liver function. It does not necessarily predict worse effects to come (in a given person). It is not a valid measure of severity of liver injury or dysfunction. It is too unspecific to be reliable in screening for relatively rare effects on the liver. Although these are substantial limitations, ALT is a very useful biomarker if understood and used properly. It is important to consider how and why these erroneous concepts came to have such wide acceptance, and how elevations of ALT activity for evaluating patients and subjects under study might be interpreted better.

A small set of serum chemical tests for liver injury or dysfunction is now considered "classic," namely: the serum concentration of total bilirubin (TBL) and the serum activities of alkaline phosphatase (ALP), aspartate aminotransferase (AST), and ALT.

Of these tests, the elevated serum enzyme activities may indicate injury to hepatocytes or biliary cells, but only the bilirubin elevation indicates any loss of liver function. These tests have been "validated" and "qualified," not by opinions of an expert committee, but by clinical use for more than six decades of successful application in millions of cases. Serum TBL was initially used in 1913 to distinguish hemolytic jaundice from hepatic jaundice,[1] and ALP was initially used in 1930 to distinguish cholestatic jaundice from hepatocellular jaundice.[2] Assay of the two serum transaminases (AST, ALT) was utilized by cardiologists to study postoperative myocardial infarction by detecting heart injury, not liver injury.[3] Their search for new biomarkers of such injury started in 1952, and they first reported[4] in 1954 the use of serum glutamic oxalacetic transaminase (SGOT) as a marker to distinguish between acute myocardial infarction and other heart diseases, infections, neoplastic diseases, and 50 normal controls. The full paper[5] in 1955 detailed a chromatographic method to assay glutamate produced by the transamination, and a brief note described a new, rapid spectrophotometric method devised by a medical student working with the team.[6]

The transamination reaction occurs not just in the liver but also in many other tissues, exchanging an amino-group for a keto-group, interconnecting protein and carbohydrate metabolism, and showing activity in liver cytoplasm and mitochondria, cardiac and skeletal muscle, intestinal epithelial cells, kidney, and many other tissues (**Figure 1**).[7] An important cofactor is pyridoxal phosphate,[8] derived from pyridoxine (vitamin B_6). The reaction is reversible, but the removal of oxalacetate by reduction to malate using malic dehydrogenase moves it steadily to the right. The second reaction required[8] reduced nicotinamide adenine dinucleotide (NADH) or diphosphopyridine nucleotide. The disappearance of reduced NADH at ultraviolet wavelength 340 nm during spectrophotometric analysis could be followed easily, and the activity of SGOT was expressed as Karmen units, a decrease of 0.001 in optical density per minute (after reading the Karmen description of the new rapid spectrophotometric method, the author, then a medical intern, obtained permission to use the clinical laboratory of the Hospital of the University of Pennsylvania at night and ran several hundred tests of serum AST and ALT on his and other patients, until the hospital added it to its list of routine tests). Using the Horecker-Kornberg molar extinction coefficient of 6.22×10^{-6} cm^2 per mol,[9] the change in optical density can be converted to µmol/min/ml of serum, allowing consistent comparisons to be made with other methods. The peak value of SGOT (AST), when measured serially, was found to be approximately proportional to the extent of myocardial infarction.[10]

[1]Office of Pharmacovigilance and Epidemiology, Center for Drug Evaluation and Research, US Food and Drug Administration, Federal Research Center at White Oak, Silver Spring, Maryland, USA. Correspondence: JR Senior (john.senior@fda.hhs.gov)

Received 8 May 2012; accepted 8 June 2012; advance online publication 1 August 2012. doi:10.1038/clpt.2012.108

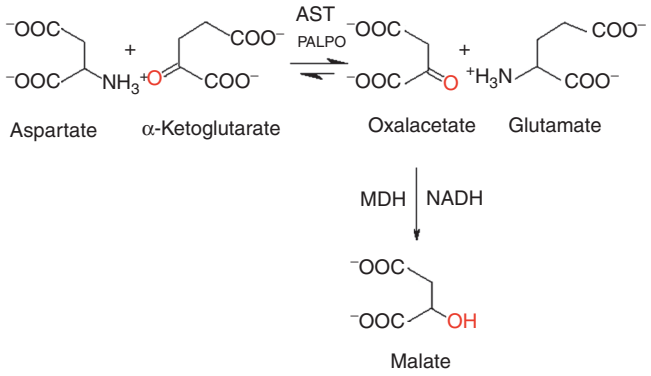

Figure 1 Interaction between amino acids and keto acids during transamination. AST, aspartate aminotransferase; MDH, malic dehydrogenase; NADH, nicotine adenine dinucleotide; PALPO, pyridoxal phosphate.

Another enzyme that was found in heart and skeletal muscle, but that was even more concentrated in liver cells, was serum glutamic pyruvic transaminase (SGPT), which catalyzed transfer of the amino-group from alanine to α-ketoglutarate, giving rise later to the name ALT (sometimes also referred to as ALAT). That reaction was also "pulled" to the right by adding lactate dehydrogenase to speed the reduction of pyruvate, with spectrophotometric measurement of NADH disappearance at wavelength of 340 mμ. The new, rapid, simple, and inexpensive method for determining the activity levels of serum enzymes such as AST, ALT, and others, revolutionized clinical chemistry, and led to the gradual abandonment of the poorly understood and unspecific cephalin flocculation and thymol turbidity tests of the late 1950s and early 1960s. It was quickly recognized that SGOT elevation was a measure of liver cell injury, but not of liver function,[11] and confirmed[12–14] that SGPT was usually more elevated in liver injury than SGOT was, although many laboratories continued to assay both. The two transaminases were initially named for the products formed, namely, SGOT and SGPT. However, after the revision of the confusing nomenclature by the International Union of Biochemistry in 1961, these were redesignated as AST and ALT, respectively, to indicate the substrate amino acids aspartate and alanine.[15] Spectrophotometric assays for AST and ALT activities could be automated to provide quick analyses. Currently, we express a unit (U) of serum enzyme activity in terms of micromoles per minute of substrate converted. However, the Système International is also used, mainly in Europe, and in that system the unit of enzyme activity is the katal (kat), expressed in moles per second, calling for conversion (1,000 nmol per μmol/60 s per minute =16.67, therefore U = 16.67 nkat).

However, methods used for the determination of ALT activity have not been standardized, and a definition of "normal" has not been agreed upon; it varies from laboratory to laboratory and depends on which population group is defined as "normal." Consequently, caution is needed in the interpretation of results obtained using varying substrates, methods, conditions, and instruments, and different definitions of what the "normal" ranges are.

Other simpler, colorimetric methods[16,17] were also proposed, based on the 1950 finding[18] of the chemical reduction of oxalacetate to pyruvate by aniline citrate, and color formation with addition of dinitrophenylhydrazone. However, the automated spectrophotometric method is currently the one that is most used.

It is unclear when the term "liver function test" was first used to refer to the determination of ALT and AST values, but they were included[19] among other tests in 1958. SGOT (AST) was useful initially, but was found to be nonspecific for heart muscle, and has subsequently been supplanted by more specific biomarkers such as cardiac troponin. Serum ALT activity became the leading biomarker for acute liver injury or disease; although AST is often measured as well, this value is usually redundant.

A giant step forward came with the appreciation that hepatocellular injury, if sufficiently extensive as to cause overall liver dysfunction with jaundice, was a lesion associated with a substantial rate of mortality if caused by a drug or chemical. This was first stated by Zimmerman[20] in 1968, based on his vast clinical experience with thousands of cases, and was restated in 1978 in his classic text,[21] *Hepatotoxicity, the Adverse Effects of Drugs and Other Chemicals on the Liver*. At a Fogarty International Conference[22] at the National Institutes of Health (NIH) in 1974, and at a second conference[23] in 1978, attempts had been made at reaching consensus on the nomenclature relating to liver disease and to agree upon a standard for interpreting laboratory findings suggestive of hepatotoxicity caused by drugs and chemicals. There was consensus that elevation of ALT to more than three times the upper limit of the normal range (>3× ULN) was "markedly abnormal." At that meeting, Robert Temple of the US Food and Drug Administration (FDA) was impressed by Zimmerman's observation that "drug-induced hepatocellular jaundice is a serious lesion"; despite the omission of that idea in the published proceedings, Temple began applying it to reports of studies submitted to the FDA.

In 1982, the National Cancer Institute (NCI) at the NIH began to grade symptoms, clinical findings, and clinical laboratory tests and findings into categories of severity of abnormality, designated as the Common Toxicity Criteria (CTC).[24] These grades have been carried forward to the present and are now widely used by oncologists and many other physicians. The grades were set at fixed, consistent levels for all serum enzyme increases, with cut-off levels of activities at >2.5, >5, and >20, as separating mild, moderate, severe, and life-threatening degrees of elevation, all expressed in multiples of the upper limit of the normal range (×ULN, for whichever laboratory was performing the assay). Included among serum enzymes were ALT, AST, ALP, and others such as γ-glutamyl-transferase (GGT), and creatine phosphokinase. It was reported that the fixed levels of elevation suggested for the CTC were established by consensus among experts consulting to the NCI and not from extensive analyses of clinical data.

In arriving at the criteria for diagnosing drug-induced liver disorders by distinguishing between primary hepatocellular injury and biliary obstructive processes, there was consensus[25] that relative levels of ALP elevation could be used as markers to indicate cholestatic problems, whereas elevated ALT or AST levels would indicate hepatocellular injury. It was suggested that

the ratio of ALT to ALP, referred to as *R*, could be used at the time of recognition of liver injury to determine the nature of the injury: cholestasis if $R < 2$, primary hepatocellular injury if $R > 5$, and mixed if $2 < R < 5$.

A primary difficulty was in distinguishing drug-induced liver injury (DILI) from that caused by disease processes such as viral hepatitis, biliary tract disorders, acute Wilson's disease, acute alcoholic hepatitis, autoimmune hepatitis, Gilbert syndrome, congenital abnormalities, and other non-drug-related causes. It was found that DILI was extremely variable in its presentation, tending to mimic known liver diseases; there were no pathognomonic findings associated with it, even from a liver biopsy. As a result, DILI was considered a diagnosis of exclusion after all other causes had been ruled out as being very unlikely. The group in France was expanded with the addition of international consultants. This group created a scoring system known as RUCAM[26,27] (Roussel-Uclaf Causality Assessment Method). It used key points of evidence to yield an estimate of the likelihood of DILI. The reasoning underlying the development of this method owed much to the thoughtful earlier work of Hill,[28] Irey,[29] and Naranjo[30] on the appropriate information for causal attributions. The RUCAM approach was extensively used to standardize the clinical estimation of causal attribution, followed by a modification[31] of the NCI common toxicity criteria in 1998 by the inclusion of a separate scale for attribution of causality of adverse events.

Despite the proposal[26,27] of a structured scale to evaluate the probability or likelihood of causal association between a drug or experimental agent and an adverse effect on the liver, and subsequent modifications,[32,33] the assignment of a number representing the likelihood that it was the drug that had caused the adverse effect still required some medical judgment and opinion that was very difficult to standardize. After the DILI Network was set up by the NIH in 2003, the members found difficulty in agreeing[34] on levels of causal likelihood using RUCAM, and went back to individual case assessments.

Meanwhile, the FDA recognized that there was a requirement for a more consistent approach to determining whether a new drug under evaluation for possible approval for clinical use in patients and marketing was causing liver injury.[35] The FDA (Temple) definition of Zimmerman's observation was modified as non-cholestatic elevation of ALT >3× ULN and TBL >2× ULN if no alternative cause could be found, as applied to data gathered from controlled clinical trials, followed in 2000 by an NIH conference[36] to discuss mechanisms of DILI. In 2001 the FDA, together with the American Association for the Study of Liver Diseases and the Pharmaceutical Research and Manufacturers of America, initiated a series of annual meetings on DILI that has continued to the present; the programs, slides, and commentaries are accessible[37] on the Internet. In 2008, the DILI Network was expanded from five to eight sites in the United States. In 2007, the FDA published a draft guidance for industry on the criteria for assessing DILI in clinical trial data and invited comments. The final[38] version of the guidance was issued in 2009 after consideration of the comments received. The guidance document retained 3× ULN of ALT as an initial warning level, with 5× ULN indicating more serious level, and 8× ULN being an indication to stop administration of the investigative drug.

COMBINED ALT AND TBL MEASUREMENTS: A NOT-SO-NEW BIOMARKER

The principle elucidated by Hy Zimmerman was employed in a software project devised at the Center for Drug Evaluation and Research (CDER) under its Regulatory Science Review enhancement program to assist medical reviewers in finding cases of special interest and potential concern among thousands of subjects in large clinical trials. The research program, named "eDISH" (Evaluation of Drug-Induced Serious Hepatotoxicity), was developed in 2004–2009 with modest support from the Center for Drug Evaluation and Research Regulatory Science Review program. Using standardized data entry for the serial measurements of the serum ALT and TBL carried out during large controlled clinical trials, the program displayed the highest observed values of ALT for each subject on the abscissa of the graph and the highest observed TBL value on the ordinate to produce a single point, ALT and TBL plotted as the $\log_{10}$ values of >×ULN (so as to keep the much more widely variable ALT values within the same graphical range as the less variable TBL values, while preserving evidence of changes in these values). The arbitrarily selected cutoff levels (ALT >3× ULN and TBL >2× ULN) are not necessarily fixed; they were conservatively set to include cases that might be of interest. The horizontal and vertical lines reflecting those levels divide the graph into four quadrants: lower left, in or near the normal range; upper left (TBL elevations without ALT rises of note), representing possible cholestatic problems; lower right (ALT elevations without TBL elevations >2× ULN), "Temple's Corollary," referred to as such because of Temple's observation that more serious cases usually arose out of a background of increased incidence of mild elevations in ALT levels in subjects on the experimental drug as compared with controls; and upper right, the indicators of possible Hy's Law cases, depending on whether the findings were of hepatocellular rather than cholestatic type, and whether no cause other than drug-induced could be found.

The computer search of subjects' data looked first for maximum ALT and TBL elevations for hyperbilirubinemia (>2× ULN: less than the level at which there is yellowing of skin or sclerae) over the period of study observation, mainly for hepatocellular injury of a mild degree as opposed to initially cholestatic processes. The cutoff levels of 3× ULN for ALT and 2× ULN for TBL were taken from the FDA guidance document.[38] These levels are conservatively set so as to include cases for closer inspection and evaluation, and are not to be considered as optimal cutoff levels. Values are plotted as logarithms (to base 10) of the actual value so as to keep the much greater fluctuations of ALT within the comparison range of the TBL fluctuations.

In the time course of changes in liver injury, markers was the second step toward assessing causality, supplemented by a third step, namely, consideration of the narrative medical history. The aim was to access the best available information so as to estimate the likelihood of drug causality of the injury and the potential severity of the outcome.

The application of this concept to studies in which thousands of subjects were observed over a period of 2 or more years, with observations recorded at several time points, allows the ready identification of subjects of highest interest, those whose data appear in the right upper quadrant (ALT >3× ULN and TBL >2× ULN), as possible Hy's Law cases (provided the rise in bilirubin levels followed the rise in ALT levels, and the narrative revealed no other cause). A display for 3,922 subjects from a single study, equally balanced between those randomized to the experimental drug (X) or to a control drug (C), is shown in **Figure 2**. It may be seen that there was a strong imbalance of ALT elevations without any rise in bilirubin (right lower quadrant, where there were 105 subjects taking X (Δ) and only 15 taking C (O)). In the right upper quadrant are the data relating to 14 subjects on X and only 1 on C. The clinical narrative data showed that the subject on C with elevations in both ALT and TBL levels had been found to have pancreatic carcinoma. Of the 14 subjects on X, only in ~50% were there any possible alternative explanations for the findings; the others presumably represent injury and dysfunction induced by the experimental drug X.

With this approach, it is possible to point to/click on individual symbols (subjects) on the computer screen and obtain a graphic display of all the data for that individual subject over his/her entire observation time so as to visualize the time-related changes in values, including those of ALP and AST (**Figure 3**). These changes give physicians the initial information to judge causality. An eDISH plot of the time course of all liver test values (ALT, AST, ALP, and TBL) for a selected subject (chosen by clicking on a symbol on the x–y plot of the log-log values) allows the second step of the eDISH evaluation, as shown by the time course of ALT, AST, ALP, and TBL for a single subject (O) in the right upper quadrant in the x–y plot of the 3,922 subjects of the study.

Note that the experimental drug (control: warfarin) was stopped after TBL had already risen, with increase in ALP, to levels relatively greater than those of either ALT or AST (**Figure 4**). The time course information was then supplemented in the third step of the eDISH program with a medical narrative to reveal that the selected subject had pancreatic carcinoma. The narrative (available by clicking on a link on the time course graph), is meant to permit a physician knowledgeable about the process and art of differential diagnosis to assess the most probable cause of the abnormalities shown, as well as the severity of the clinical problem.

The power of the computer to process a very large amount of data from hundreds or thousands of subjects who had been observed serially over 2 years was thereby combined with the facility of rapid pattern recognition at a glance by a physician. The eDISH program has now been used to help medical reviewers select, from very large clinical trials in studies of more than 18,000 subjects under single protocols, the relatively small numbers of cases of serious liver toxicity that might be caused by drugs for the purpose of evaluating the safety of new drugs under consideration for approval. Serious cases of DILI are generally quite rare; narratives or detailed diagnostic assessment are frequently not necessary. The eDISH program has been found to be of use to academic centers, to pharmaceutical companies seeking approval of new

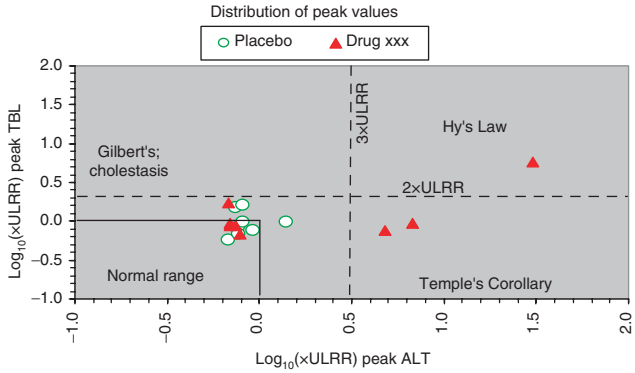

Figure 2 Application of the Zimmerman concept in graphical format. ALT, alanine aminotransferase; TBL, total bilirubin.

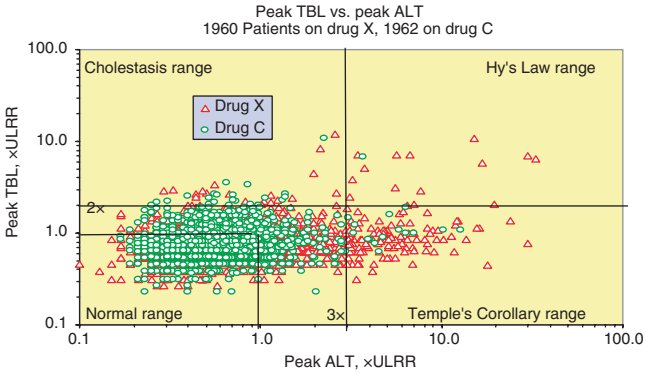

Figure 3 Evaluation of Drug-Induced Serious Hepatotoxicity (eDISH), applied to a study of 3,922 subjects observed for 2 years (eDISH step 1). ALT, alanine aminotransferase; TBL, total bilirubin.

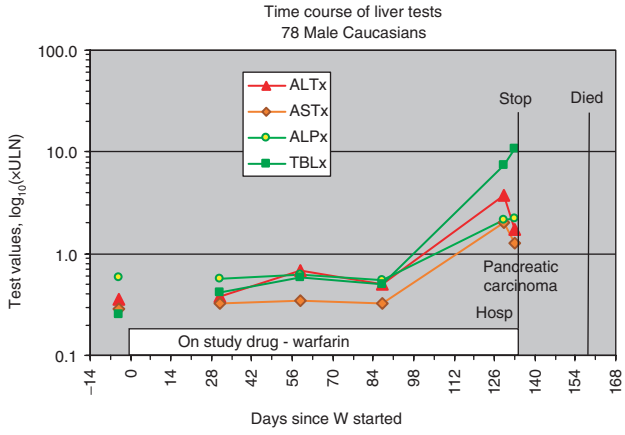

Figure 4 Time course of serial key liver tests in a selected patient (research subject) observed for ~5 months. ALP, alkaline phosphatase; ALT, alanine aminotransferase; AST, aspartate aminotransferase; TBL, total bilirubin; ULN, upper limit of the normal range.

products through confidential analyses before data submission,[39] and as a research tool for evaluating large datasets of information relating to liver injury and dysfunction. The eDISH program has been used at the FDA by medical reviewers and consultants to assist in reviewing the data of >30,000 subjects receiving a single new drug, and also for more than 100,000 subjects in phase III clinical trials, receiving various experimental or control drugs.

OTHER TESTS

Other tests to identify liver disorders have been proposed and are favored by some. However, these have not been adopted for widespread clinical use. Some of the tests were found to be redundant, but the main problem was insufficient specificity, leading to false-positive results. This lack of specificity was particularly problematic when the disorder was of rare incidence or prevalence, making the diagnostic value of a positive result very low indeed. The measurement of ALT levels has been criticized as being insufficiently specific to the detection of liver injury. Among the tests still used by some physicians are those relating to other serum enzymes: γ-glutamyl-transferase, lactate dehydrogenase, isocitric dehydrogenase, 5′-nucleotidase, leucine aminopeptidase, glutathione-sulfotransferase, cholinesterase, glutamate dehydrogenase, and others. Other physicians also continue to study the concentrations of bile acids, serum albumin, and various globulins. Another category of tests includes measures of the clearance of injected dyes such as bromosulfophthalein and indocyanine green, galactose clearance, the synthesis rate of urea, oxidative metabolism of various substrates, and collection of expired radioactive or isotopic carbon dioxide. In general, these tests are low in sensitivity and specificity, or cumbersome, time-consuming, laborious, and expensive.

Some other serum enzymes have been proposed[40] as "new" biomarkers that might be more specific and perhaps even more sensitive than ALT. These include paraoxonase; malate, sorbitol, or glutamate dehydrogenase; and purine nucleoside phosphorylase, as commercially available and readily measurable using photometric methods. Tests for serum F protein, arginase I, and glutathione-sulfotransferase-α are considered to be expensive. However, none of the proposed new biomarkers has been evaluated against the not-so-new combination biomarker of (ALT and TBL) that is both quite sensitive in detecting cases and highly specific for liver problems. There is great need for new biomarkers[41] that are highly specific for acute serious hepatocellular DILI and truly predictive of outcome. This need is now widely recognized, resulting in a deluge of recent publications.

Another issue to be resolved is whether the comparator level used to estimate elevation should be the normal range of levels in a healthy population, or the pretreatment (if only one) or "baseline" (if two or more) values of that particular subject. Data for such assessments can best be obtained from controlled clinical trials, and this represents a good research project that has yet to be undertaken adequately.

SENSITIVITY AND SPECIFICITY

When both sensitivity and specificity are considered together to evaluate a binary test of positive or negative results, it has become the norm to plot (1—specificity; false positive) on the abscissa and sensitivity (true positive) on the ordinate as a receiver operating characteristic curve. This procedure was developed by electrical engineers seeking the best cutoff values for radar signals from enemy planes, a scenario wherein errors are costly, whether false negative (incoming enemy planes may be missed) or false positive (unnecessary launch of defense aircraft). This concept was subsequently applied in the field of psychology and then more widely in medicine.[42] Most diagnostic tests are imperfect, and it is therefore useful to find the cutoff value that will yield optimal results.

Figure 5 shows a hypothetical but realistic receiver operating characteristic curve for ALT values, showing various cutoff levels. It is obvious that there is a trade-off between sensitivity and specificity, a gain for one representing a loss for the other. It is also apparent that there is a relatively large loss of specificity when the sensitivity cutoff is reduced from 3 to 2.5× ULN, but little gain in specificity when sensitivity is raised from 3× ULN to 5, 8, or even 10× ULN. The best cutoff value is taken as the point on the curve most distant from the diagonal line of null diagnostic value. Specificity is the elusive and more important measure, especially if relatively rare events are being sought, as is usually the case when searching for serious hepatotoxic events caused by drugs. At the 1978 Fogarty[23] meeting, hepatology experts chose 3× ULN as the cutoff for ALT and other serum enzyme activity levels, based on expert consensus and not on data analyses. The determination of the best cutoff point on the curve should preferably be based on clinical values and tolerance of α (false-negative) and β (false-positive) errors, now that vast amounts of real data are becoming available. For rare events, as serious DILIs usually are, false-positive errors are more troublesome, and therefore extremely high specificity is required to offset their effects. It is often forgotten or misunderstood that positive "predictive" values of a test (true-positive test results/all positive test results)[43](it is perhaps unfortunate that Vecchio[43] coined the adjective "predictive" for values of test results—to me, it seems that "diagnostic" would have been better, but the term "predictive" is now deeply entrenched) are very highly dependent on the rate of incidence of the result sought, falling off very sharply for relatively uncommon or rare events.[44] Even a very good test with 0.95 sensitivity and 0.90 specificity will show correct positive results in 90.5% of those tested when the incidence

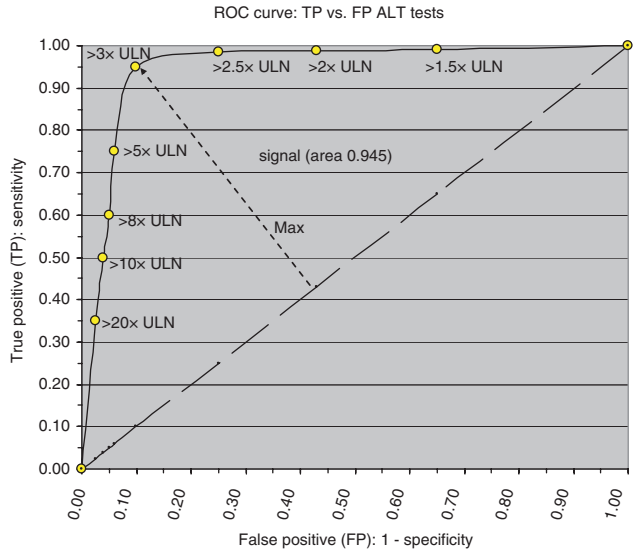

Figure 5 Receiver operating characteristic (ROC) curve for various cutoff points determining positive or negative test results for serum alanine aminotransferase (ALT) activity levels. ULN, upper limit of the normal range.

rate is 50%, but only in <0.01% when the incidence is 1 per 1,000 tested. At such low incidence rates, extremely high test specificity, on the order of 99.9%, is needed to give even a 50% chance that a positive test result is true. One should be cautious when interpreting published reports of high sensitivity for negative results, because these are expected and very likely for rare events, and are therefore not useful in this context (**Figure 6**).

WHAT DOES ALL THIS ADD UP TO?

Since the 1978 Fogarty Conference and 1982 NCI toxicity criteria proposal, vast amounts of data have been accumulated. It is now obvious that elevations in serum enzyme levels alone are not necessarily measures of clinical severity. Elevated levels of ALT do not predict what will happen; they only indicate what has happened up to the time of measurement. Serum ALT enzyme activity levels are not measures of liver function. It is not a function of the liver to regulate the levels of activity of serum enzymes that have no function in the plasma but represent leakage or release from damaged cells in which they previously did have functions. In plasma, the levels of activity are determined by the rates of release from cells and by their rates of inactivation or degradation by nonspecific proteases. To a greater extent the TBL or prothrombin time (or derived international normalized ratio) are better measures of the ability of the whole liver to perform its true functions, such as clearing the plasma of bilirubin or synthesizing important coagulation proteins.

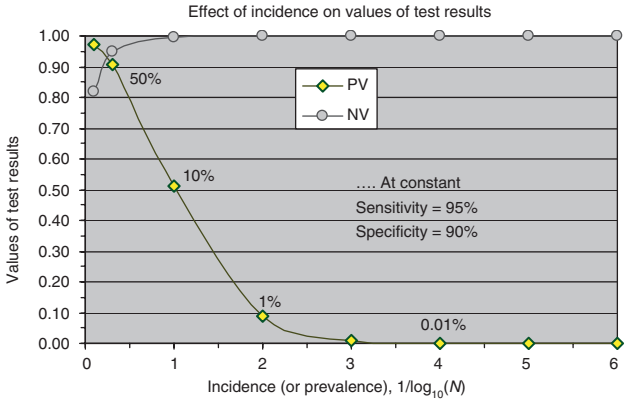

Figure 6 Drastic loss of diagnostic value for positive test results when incidence or prevalence of the event is very low. NV, value of negative test result; PV, value of positive test result.

I now propose that the gradings of severity (**Table 1**) published since 1982 by the NCI as the Common Toxicity Criteria for Adverse Events be considered instead as measures of urgency. To do what?—(i) to repeat the tests immediately (in local laboratories to avoid the delay involved in obtaining results when specimens are sent to distant central laboratories); (ii) to confirm the findings, and establish the direction and rapidity of change; (iii) to initiate additional studies to clarify the probable cause (**Table 2**) of the findings if they become worrisome; (iv) to preserve serum samples for possible retrospective tests to be done later as needed.

Elevations in the concentration levels of serum TBL, or of the direct-reacting, conjugated bilirubin, are highly specific to liver problems but are not very sensitive as markers; also, the elevation often occurs only late in a disease process. The search for new biomarkers should accept the combined biomarker of (ALT and TBL) as the standard to surpass, rather than ALT alone. Under the new regulations of 2010,[45,46] it is the responsibility of the sponsors of drug trials to supervise local investigators and guide them on the procedure to follow when abnormal liver test results are found; all data are required to be recorded in the case report forms, both in local and central laboratories.

Despite the limitations on the predictive value of ALT elevations in individual subjects, it has been found repeatedly that the increased occurrence of such elevations in groups that are being studied is highly predictive of what is later likely to be found in other groups. This is an important point, the significance of which is often misunderstood.[47–49]

RECOMMENDATIONS

Proposed here are some new interpretations relating to elevated levels of ALT activity, and suggestions for how they might lend themselves to better general use, to trigger discussions, objections, or corrections, to stimulate debate, and to aim at consensus and more consistent application:

1. The measurements of serum ALT (as well as AST and ALP) are not tests of liver function, but of liver cell injury.
2. Whole-liver dysfunction is what really determines the severity of the injury.
3. Serum bilirubin concentration does measure at least one function of the liver, namely, to clear plasma of bilirubin, and depends on whole-liver functional capacity.

Table 1 NCI common toxicity criteria grades of severity, 1982–2012

Grade test	1 (Mild)	2 (Moderate)	3 (Severe)	4 (Life-threatening)	5 (Fatal)[a]
ALT, U/ml	>1–2.5 ×ULN	>2.5–5 ×ULN	>5–20 ×ULN	>20 ×ULN	—
ALP, U/ml	>1–2.5 ×ULN	>2.5–5 ×ULN	>5–20 ×ULN	>20 ×ULN	—
AST, U/ml	>1–2.5 ×ULN	>2.5–5 ×ULN	>5–20 ×ULN	>20 ×ULN	—
CPK, U/ml	>1–2.5 ×ULN	>2.5–5 ×ULN	>5–20 ×ULN	>20 ×ULN	—
GGT, U/ml	>1–2.5 ×ULN	>2.5–5 ×ULN	>5–20 ×ULN	>20 ×ULN	—
TBL, mg/dl	>1–1.5 ×ULN	>1.5–3 ×ULN	>3–10 ×ULN	>10 ×ULN	—

ALP, alkaline phosphatase; ALT, alanine aminotransferase; AST, aspartate aminotransferase; CPK, creatine phosphokinase; GGT, γ-glutamyl-transferase; NCI, National Cancer Institute; TBL, total bilirubin; ×ULN, multiples of the upper limit of normal.

[a]Or liver transplantation.

Table 2 NCI attribution of causality of adverse events

Code	Description	Definition
5	Definite	Clearly related to the investigational agents(s)
4	Probable	Likely related to the investigational agents(s)
3	Possible	May be related to the investigational agents(s)
2	Unlikely	Doubtfully related to the investigational agents(s)
1	Unrelated	Clearly not related to the investigational agents(s)

NCI Common Toxicity Criteria version 2, 1999.

NCI, National Cancer Institute.

4. The combined biomarker (ALT and TBL) has the high sensitivity of ALT and the very great specificity of TBL, and is the current standard to surpass for proposed new biomarkers.
5. Urgency (to repeat the test and initiate special study) should replace severity in the NCI CTC guideline relating to serum enzyme activity levels.
6. Elevated ALT activity does not necessarily predict what will happen to the liver for an individual but is of predictive value when found in groups.
7. There is need for standardization, both of the methods used to measure ALT activity and of what should be adopted as the truly "normal" range.[50,51]
8. Suitable revisions in guidances, teaching approaches, and clinical practice will need to reflect these points.

The use of correct terminology is not just a trivial pedantic exercise but reflects how physicians think and act. The application of these points may have consequences on weighty sources of advice such as the Common Toxicity Criteria for Adverse Events of the NCI, FDA guidances, and protocols for the conduct of clinical trials, as well as for the good practice of medicine.

ACKNOWLEDGMENTS

The discussion and recommendations are those of the author and do not reflect any official policies or positions of the US Food and Drug Administration.

CONFLICT OF INTEREST

The author declared no conflict of interest.

1. Hijmans van den Bergh, A.A. & Snapper, I. Die Farbstoffe des Blutserums. (The colored material of serum). *Deutsch Arch Klin Med.* **110**, 540–561 (1913).
2. Roberts, W.M. Variations in the phosphatase activity of the blood in disease. *Br. J. Exper. Pathol.* **11**, 90–95 (1930).
3. Wroblewski, F. & Ladue, J.S. Myocardial infarction as a post-operative complication of major surgery. *J. Am. Med. Assoc.* **150**, 1212–1216 (1952).
4. Ladue, J.S., Wroblewski, F. & Karmen, A. Serum glutamic oxaloacetic transaminase activity in human acute transmural myocardial infarction. *Science* **120**, 497–499 (1954).
5. Karmen, A., Wroblewski, F. & Ladue, J.S. Transaminase activity in human blood. *J. Clin. Invest.* **34**, 126–131 (1955).
6. Karmen, A. A note on the spectrometric assay of glutamic-oxalacetic transaminase in human blood serum. *J. Clin. Invest.* **34**, 131–133 (1955).
7. Awapara, J. & Seale, B. Distribution of transaminases in rat organs. *J. Biol. Chem.* **194**, 497–502 (1952).
8. Snell, E.E. & Jenkins, W.T. The mechanism of the transamination reaction. *J. Cell. Comp. Physiol.* **54**, 161–177 (1959).
9. Horecker, B.L. & Kornberg, A. The extinction coefficients of the reduced band of pyridine nucleotides. *J. Biol. Chem.* **175**, 385–390 (1948).
10. Agress, C.M. *et al*. Serum transaminase levels in experimental myocardial infarction. *Circulation* **11**, 711–713 (1955).
11. Wroblewski, F. & Ladue, J.S. Serum glutamic oxalacetic transaminase activity as an index of liver cell injury: a preliminary report. *Ann. Intern. Med.* **43**, 345–360 (1955).
12. Chinsky, M., Shmagranoff, G.L. & Sherry, S. Serum transaminase activity; observations in a large group of patients. *J. Lab. Clin. Med.* **47**, 108–118 (1956).
13. Ladue, J.S. & Wroblewski, F. Serum glutamic pyruvic transaminase SGP-T in hepatic disease: a preliminary report. *Ann. Intern. Med.* **45**, 801–811 (1956).
14. Chinsky, M., Wolff, R.J. & Sherry, S. Serum transaminase activity; a comparison of the pyruvic and oxalacetic transaminases. *Am. J. Med. Sci.* **233**, 400–408 (1957).
15. Report of the Commission on Enzymes of the International Union of Biochemistry (Pergamon Press, Oxford, UK, 1961).
16. De Ritis, F., Coltorti, M. & Giusti, G. Serum and liver transaminase activities in experimental virus hepatitis in mice. *Science* **124**, 32 (1956).
17. Reitman, S. & Frankel, S. A colorimetric method for the determination of serum glutamic oxalacetic and glutamic pyruvic transaminases. *Am. J. Clin. Pathol.* **28**, 56–63 (1957).
18. Tonhazy, N.E., White, N.G. & Umbreit, W.W. A rapid method for the estimation of the glutamic-aspartic transaminase in tissues and its application to radiation sickness. *Arch. Biochem.* **28**, 36–42 (1950).
19. Ragland, S. Jr. Newer liver function tests. Am. Pract. Dig. Treat. **9**, 1281–1283 (1958).
20. Zimmerman, H.J. The spectrum of hepatotoxicity. *Perspect. Biol. Med.* **12**, 135–161 (1968).
21. Zimmerman, H.J. *Hepatotoxicity: The Adverse Effects of Drugs and Other Chemicals on the Liver* 351–357 (Appleton-Century-Crofts, New York, 1978).
22. Leevy, C.M., Popper, H. & Sherlock, S. *Diseases of the Liver and Biliary Tract.* Standardization of nomenclature, diagnostic criteria and diagnostic methodology. Fogarty International Centre 1974 Proceedings No. 22, DHEW Publication No. NIH 76-725 (U.S. Government Printing Office, Washington, DC, 1975).
23. Davidson, C.S., Leevy, C.M. & Chamberlayne, E.C. *Guidelines for the Detection of Hepatotoxicity due to Drugs and Chemicals (1978).* U.S. Department of Health, Education, and Welfare, Public Health Service, National Institutes of Health. NIH Publication 79-313 (1979).
24. National Cancer Institute Common Toxicity Criteria, Version 1, 1982 <http://www.ucdmc.ucdavis.edu/clinicaltrials/StudyTools/Documents/NCI_Toxicity_Table.pdf>.
25. Bénichou, C. Criteria of drug-induced liver disorders. Report of an international consensus meeting. *J. Hepatol.* **11**, 272–276 (1990).
26. Danan, G. & Benichou, C. Causality assessment of adverse reactions to drugs–I. A novel method based on the conclusions of international consensus meetings: application to drug-induced liver injuries. *J. Clin. Epidemiol.* **46**, 1323–1330 (1993).
27. Benichou, C., Danan, G. & Flahault, A. Causality assessment of adverse reactions to drugs–II. An original model for validation of drug causality assessment methods: case reports with positive rechallenge. *J. Clin. Epidemiol.* **46**, 1331–1336 (1993).
28. Hill, A.B. The environment and disease: association or causation? *Proc. R. Soc. Med.* **58**, 295–300 (1965).
29. Irey, N.S. Diagnostic problems in drug-induced diseases. *Ann. Clin. Lab. Sci.* **6**, 272–277 (1976).
30. Naranjo, C.A. *et al*. A method for estimating the probability of adverse drug reactions. *Clin. Pharmacol. Ther.* **30**, 239–245 (1981).
31. National Cancer Institute. Common Toxicity Criteria, version 2, 1999 <http://ctep.cancer.gov/protocolDevelopment/electronic_applications/docs/ctcmanual_v4_10-4-99.pdf>.
32. Maria, V.A. & Victorino, R.M. Development and validation of a clinical scale for the diagnosis of drug-induced hepatitis. *Hepatology* **26**, 664–669 (1997).
33. Aithal, G.P., Rawlins, M.D. & Day, C.P. Clinical diagnostic scale: a useful tool in the evaluation of suspected hepatotoxic adverse drug reactions. *J. Hepatol.* **33**, 949–952 (2000).
34. Rockey, D.C. *et al.*; US Drug-Induced Liver Injury Network. Causality assessment in drug-induced liver injury using a structured expert opinion process: comparison to the Roussel-Uclaf causality assessment method. *Hepatology* **51**, 2117–2126 (2010).
35. FDA Conference. Drugs and the liver: what they do to each other. April, November 1999 <http://www.aasld.org/dili/Pages/default.aspx>.
36. Bissell, D.M., Gores, G.J., Laskin, D.L. & Hoofnagle, J.H. Drug-induced liver injury: mechanisms and test systems. *Hepatology* **33**, 1009–1013 (2001).

37. American Association for the Study of Liver Diseases: Education/Training, Drug-Induced Liver Injury 2012 Program<http://www.aasld.org/education/Pages/2012HepatotoxicitySpecialInterestGroup.aspx >.
38. US Food and Drug Administration. Guidance for industry on drug-induced liver injury: premarketing clinical evaluation. *Fed. Regist.* **74**, 38035–38036 (2009) (notice of availability). Full document see <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM174090.pdf>.
39. Watkins, P.B. *et al.* Evaluation of drug-induced serious hepatotoxicity (eDISH): application of this data organization approach to phase III clinical trials of rivaroxaban after total hip or knee replacement surgery. *Drug Saf.* **34**, 243–252 (2011).
40. Ozer, J.S. *et al*. Recommendations to qualify biomarker candidates of drug-induced liver injury. *Biomark. Med.* **4**, 475–483 (2010).
41. Starkey Lewis, P.J. *et al*. Circulating microRNAs as potential markers of human drug-induced liver injury. *Hepatology* **54**, 1767–1776 (2011).
42. Ledley, R.S. & Lusted, L.B. Reasoning foundations of medical diagnosis; symbolic logic, probability, and value theory aid our understanding of how physicians reason. *Science* **130**, 9–21 (1959).
43. Vecchio, T.J. Predictive value of a single diagnostic test in unselected populations. *N. Engl. J. Med.* **274**, 1171–1173 (1966).
44. Krieg, A.F., Gambino, R. & Galen, R.S. Why are clinical laboratory tests performed? When are they valid? *JAMA* **233**, 76–78 (1975).
45. FDA. Investigational new drug safety reporting requirements for human drug and biological products and safety reporting requirements for bioavailability and bioequivalence studies in humans: final rule. *Fed. Regist.* **75**, 59935–59963 (2010).
46. Sherman, R.B., Woodcock, J., Norden, J., Grandinetti, C. & Temple, R.J. New FDA regulation to improve safety reporting in clinical trials. *N. Engl. J. Med.* **365**, 3–5 (2011).
47. Kaplowitz, N. Idiosyncratic drug hepatotoxicity. *Nat. Rev. Drug Discov.* **4**, 489–499 (2005).
48. Weil, J.G., Bains, C., Linke, A., Clark, D.W., Stirnadel, H.A. & Hunt, C.M. Background incidence of liver chemistry abnormalities in a clinical trial population without underlying liver disease. *Regul. Toxicol. Pharmacol.* **52**, 85–88 (2008).
49. Llanos, L. *et al*. The existence of a relationship between increased serum alanine aminotransferase levels detected in premarketing clinical trials and postmarketing published hepatotoxicity case reports. *Aliment. Pharmacol. Ther.* **31**, 1337–1345 (2010).
50. Dufour, D.R. Alanine aminotransferase: is it healthy to be "normal"? *Hepatology* **50**, 1699–1701 (2009).
51. Ruhl, C.E. & Everhart, J.E. Upper limits of normal for alanine aminotransferase activity in the United States population. *Hepatology* **55**, 447–454 (2012).