CDER-PhRMA-AASLD Conference 2000

Clinical White Paper
November 2000

I. Objective:

To find clinical development and pre-approval review strategies that will detect drugs with serious hepatotoxic potential, without unnecessarily identifying non-hepatotoxic drugs as potentially toxic.

II. Background:

Through the years serious drug-induced hepatotoxicity has been the most common single unexpected or "idiosyncratic" adverse effect that has had major effects on drug marketing (preventing marketing or causing withdrawal) and clinical use (limiting a drug to second-line status, requiring special monitoring or restricted use). Serious hepatotoxicity refers here to liver injury (usually, but not always, acute hepatocellular necrosis) that leads to liver failure and death or need for liver transplant. In general, the hepatotoxicity of interest is "idiosyncratic," meaning not predictable or dose-related, in contrast to such drugs as chloroform and carbon tetrachloride, but that distinction may need a critical look. "Idiosyncratic" toxicity may be dose-related in people with unusual metabolism, or in particular circumstances. Certainly acetaminophen toxicity, which is mechanistically thought to be similar to isonizid/iproniazid toxicity, is dose-related. In general, cholestatic effects are of interest but not life-threatening.

Examples of drugs withdrawn because of hepatotoxicity include:

MARSILID (iproniazid)	1959	DURACT (bromfenac)	1998
SELECRYN (ticrynafen)	1979	REZULIN (troglitizone)	2000
ORAFLEX (benoxaprofen)	1982		

Examples of drugs never marketed in the United States include:

Ibufenac	1965	Tasosartan	1998
Perhexilene	1976	Fialuridine (FIAU)	1996
Dilevalol	1990		

Examples of drugs with significant limitations of use (warnings, dose restrictions, monitoring) include:

NIACIN (nicotinic acid)	1961	CYLERT (pemoline)	1995
Isoniazid	1969	FELBATOL (felbamate)	1997
DANTRIUM (dantrolene)	1976	ZYFLO (zileutan)	1997
TYLENOL (acetaminophen)	1977	TASMAR (tolcapone)	1998
NORMADYNE (labetalol)	1989	TROVAN (trovafloxacin)	1999

The types of liver injury leading to limitations of use have been diverse, including a cirrhotic picture (perhexilene), cholestasis (benoxaprofen), and even what appears to have been primary mitochondrial injury (FIAU), but most have involved hepatocellular injury.

The drugs cited above are conspicuous because hepatotoxicity was recognized late, only after marketing in the U.S. or elsewhere (except for FIAU and tasosartan). Many other drugs have undoubtedly been discarded because of animal or early human evidence of liver toxicity, but there has never been a systematic accounting of these cases and it is not known when in development of these drugs that this occurred, or what the basis for the decision was. It is obviously most desirable that the hepatoxic potential of drugs be discovered as early as possible, using effective pre-clinical screening and sensitive and specific human testing. If toxicity is nonetheless not

detected pre-marketing, we need to discover it as rapidly as possible after marketing, if possible with a reasonable estimate of risk. Other groups will develop discussions of pre-clinical and post-marketing procedures. The purpose of this paper is to consider how we can enhance detection of hepatotoxic potential during clinical development and pre-approval clinical review of drugs.

III.   Clinical Evaluation of Hepatotoxic Potential:

Success in detecting hepatotoxic potential and making good decisions based on the evidence depends on:

- Conducting the best tests for liver abnormalities during clinical development and doing them at the most appropriate frequency.
- Interpreting the findings optimally; i.e., identifying "signals" of hepatotoxicity and interpreting them based on knowledge of the performance characteristics (sensitivity, specificity, predictive value) of those signals.
- Considering appropriateness of particular actions in various situations, such as labeling for potential risk, encouraging monitoring at particular intervals, or seeking a larger database prior to approval to evaluate the signal further. Much of this decision-making depends on knowing the particular characteristics of various signals.

At least so far, it appears that drug manufacturers carry out appropriate liver testing during clinical development. Better use of these data depends on better understanding of signals of hepatotoxic potential. The focus of this paper will therefore be on: 1) candidate signals and 2) what we can do with existing data to better describe the performance characteristics of those signals.

A. Current Testing Methods

Most sponsors collect, at regular intervals corresponding to visit frequency (generally every 2-4 weeks), serum samples from study participants for testing activities of alanine aminotransferases (ALT, also known as SGPT), aspartate aminotransferases (AST, also known as SGOT), alkaline phosphatase (ALP), and total bilirubin (Bt). They may also measure serum albumin and, much less commonly, blood prothrombin time. When important liver chemistry abnormalities emerge, they may carry out tests for hepatitis A, B, C exposure, a major advance, as in the past viral hepatitis and drug injury were difficult to distinguish, and asymptomatic or undiagnosed chronic viral hepatitis confounded interpretation of abnormal values.
It has been suggested that too frequent testing of ALT, AST, or ALP may prevent the development of greater deviations that would be informative. A case can be made for reducing the frequency of testing in at least part of the population exposed, to reflect more realistically the conditions of use that will obtain during clinical use. Frequent testing is the rule in controlled trials, but most applications have a substantial open or follow-up period in which tests are run monthly or less often. It has not been possible to discern an effect of testing frequency on the degree of abnormality seen, but it is not clear how well this has been examined.

Possibly important, although not a measure of liver injury, the metabolic pathway and the specific cytochrome P-450 (CYP450) enzymes responsible for metabolism are generally described by the time a drug is in phase 3. It is possible that certain pathways may identify a potential, or increased potential, for hepatotoxicity, but this too has not yet been established.

B.   Potential Signals; Current Interpretation and Potential Enhancement

At present, several possible signals of the ability of a drug to cause serious liver injury are regularly examined, but the performance of these "signals" is not known. The sensitivity and specificity of particular signals of hepatotoxicity potential need evaluation.

1.   Increased Rate of Abnormalities of Individual Tests, Transaminases

Overt liver failure is an obvious indicator of risk, but far more commonly, no cases are seen in an NDA database, and the lesser abnormalities seen must be interpreted. Any test result above the upper limit of the normal range (ULN) deserves notice, but slight abnormalities of this kind are common in untreated and placebo-treated patients and are not very informative about potential for serious drug injury. It has become standard practice, therefore, to look at greater deviations, such as transaminases >3x ULN, >5x ULN, etc. [Ref 1978 Fogarty conference]. Because these abnormalities (especially the milder ones) also occur spontaneously (or in patients exposed to alcohol and perhaps other influences), it is important to compare their rate in the treatment and control groups. A clearly higher rate of transaminase elevations to >3x ULN, for example, is a potential signal, as is a higher rate of more marked elevations. The comparison with a control group is probably less critical as the abnormality becomes greater (10x ULN transaminases are very unusual spontaneously) and the rate of occurrence becomes greater (3% rates of 3x ULN transaminases are not often seen in control populations). Therefore, rates of these abnormalities can be examined in the whole database, not just in the controlled trials. It seems likely, but has not been well studied, that more extreme elevations of transaminase may be better predictors of toxicity than smaller elevations, but it is possible that monitoring frequency could affect the degree of abnormality seen (frequent monitoring leads to earlier cessation of drug treatment). The rate of rise of transaminase activity might also be pertinent. It should be noted that normalization of abnormalities on continued treatment does not prove there was no signal of potential hepatotoxicity.

It is clear that some drugs (tacrine, aspirin) that cause frequent but isolated transaminase elevations, but no changes in serum bilirubin, have little or no potential for serious toxicity.

2.   Combination of Tests; Transaminase Elevation Plus Bilirubin Elevation

A combination of test results of particular interest is significant transaminase elevation (say, to at least 3x ULN) accompanied by jaundice, or perhaps more generally, by serum total bilirubin elevation, without evidence of biliary obstruction (significant elevation of ALP).

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. The explanation for this outcome was that hepatocellular injury great enough to interfere with bilirubin excretion must have involved a large fraction of the liver call mass. Recent experience at the Agency over two decades has borne out the observation that the combination of transaminase and bilirubin elevation often predicts the occurrence of severe injury in some patients. The significance of the combined abnormality was reemphasized by the late Dr. Zimmerman in the second edition of his book published posthumously in September 1999.

The idea that the combination of elevated transaminase(s) and Bt has ominous implications has come to be dubbed "Hy's Law" at the Agency. Instances (even very few of them) of transaminase elevation accompanied by elevated bilirubin (even if obvious jaundice was not present) have been associated with, and have often predicted, post-marketing serious liver injuries (fatal or requiring transplant). This has been true for bromfenac, dilevalol, troglitizone, and trovafloxacin, even though no

cases of serious injury were seen for any of these drugs pre-marketing. There do not appear to have been such cases with tolcapone, although it is still possible a larger database might have detected some. The finding of >3x ULN transaminase elevations and modest bilirubin elevation (above 2 mg /dL) in just a few patients, accompanied by an increased incidence of 3-fold transaminase elevation compared to placebo, has been the basis for refusing to approve an application. An example of this was dilevalol; after the discovery of these abnormalities in studies in the United States, European post-marketing experience was evaluated and shown to include serious liver injury. Another possible response to such a finding may be to ask for a greatly enhanced database (10-15,000 patients) prior to approval (in the case of tasosartan). It should be noted that bromfenac, troglitizone, trovafloxacin, and dilevalol had elevated rates of occurrence of >3x ULN transaminases to 2-3% in addition to several instances of elevated transaminases and bilirubin. It is possible that the detection of a few cases of combined transaminase and bilirubin elevations, together with an overall increased rate of transaminase elevations in the sample population studied, is the real "signal."

If it is generally true that a patient with elevated bilirubin and transaminase has at least a 10% chance of severe injury ("Hy's Law"), then drugs causing such liver injury will usually generate this signal at about 10 times the rate at which they cause serious liver injury. Put another way, the signal, when present, may predict a rate of serious liver injury and hepatic failure in about a tenth as many patients as showed the signal. The absence of such a signal may allow an estimate of an upper limit of the rate for serious injury. Thus, if no cases of transaminase and bilirubin elevations are seen in 3000 patients, it can be concluded that the true rate of such occurrence is not more than 1 per 1,000 (95% confidence interval: the "rule of 3"). This would then place an upper limit on the rate of serious liver injury at about 1 per 10,000.

It is not known whether assessments of other liver functions, such as albumin production or synthesis of prothrombin (measured by blood prothrombin time), would contribute to a further useful signal.

C. Evaluating Performance Characteristics (Sensitivity/Specificity) of Signals

To a degree, FDA has been operating for many years under "Hy's Law." Dilevalol and tasosartan were not approved for marketing on this basis (there was also an increased rate of transaminase elevations vs. control in these cases), but this has not been a uniform practice. Moreover, the Agency does not have a systematic evaluation of the rule's sensitivity (true positive/true positive plus false negative) or specificity (true negative/true negative plus false positive). It is possible that the specificity of Hy's Law could be enhanced (the signal of cases of combined transaminase and bilirubin elevations might show greater sensitivity and/or specificity) with modifications, such as:

- Different cut points (choice of definition of elevated transaminase or bilirubin.
- Use of information about drug metabolism.
- Analysis of time-course of events (rate of rise/fall of transaminase, bilirubin).
- Looking at cases only when there is overall increase in the rate of transaminase elevations.

We also need to know the operating characteristics of other signals, e.g., an excess of 3x, 5x, 10x, 20x elevations of transaminases (ALT and/or AST). There is need to determine whether one transaminase (perhaps ALT) is a better predictor than the other (or perhaps both ALT and AST elevated is still better).

To assess the specificity and sensitivity of potential signals and to improve the specificity and sensitivity by modifying them we will need to develop two new drug databases (lists of drugs and information about them) that will be used to describe the sensitivity and specificity of potential signals.

1. Lists

   a. Hepatotoxic Drug Data Base = List #1

A list is needed of all drugs (preferably throughout history) that have been shown to have significant hepatotoxicity, at a rate estimated to be at least 1/50,000, including drugs withdrawn or rejected for marketing in the U.S. and elsewhere, and drugs relabeled strongly (box warning, second-line status, need for monitoring) because of hepatotoxicity. The list need not include drugs that have purely cholestatic activity but should include agents where a cirrhotic or other serous injury can emerge (perhexilene, benoxaprofen). Many of the drugs on this list are known already, but by no means all are. The list includes:

isoniazid	perhexilene
iproniazid	dilevalol
dantrolene	labetalol
ticrynafen (tienilic acid)	pemoline
ibufenac	felbamate
bromfenac	tolcapone
benoxaprofen	diclofenac
zileutan	(AIDS drugs)
nicotinic acid	troglitizone
trovafloxacin	

Many more probably will be found with further searching. It would also be useful to include drugs whose development was terminated in phase 3 because of hepatotoxicity. For these "positives," the pre-marketing database will need to be reevaluated to test the performance of various signals (see 2 below), specifically, to assess their sensitivity, recognizing that failure to detect a signal in too small a database is not a failure of the signal but failure of the signal/sample size combination.

b. False Positive Signal Data Base = List #2

A list of drugs (as many as we can find) with signals of possible hepatotoxicity (increased rate of combined transaminase and bilirubin elevations of various degrees, increased incidence of transaminase elevation without bilirubin rises) that have not proved to be hepatotoxic. This is difficult because these drugs may not be identified in labeling and are not on our "list" of hepatotoxins, except where the signal is very strong, e.g., tacrine. In addition, each "signal" generates different lists. That is, the list of "Hy's Law" signals without ultimate toxicity may be much smaller than the list of elevated transaminase signals. But, of course, that is the point of this list: to evaluate the specificity of a signal.

2. Signals to be Evaluated

 The signals that should be tested are (first cut):

a) Presence of individual cases of transaminase elevation to >3x ULN with concurrent bilirubin increased by >50%, 100%. There could be separate assessments of elevated ALT, AST, or both.

b) Transaminase elevations alone. These include rates of 3x, 5x, 10x, 20x elevation compared to control or uncontrolled observations (e.g., in extension studies where there is no control group) where rates are clearly greater than background control (e.g., 3% transaminase >3x ULN). Extreme elevations of serum levels (transaminase >20x ULN) could be looked at as individual cases.

c) The combination of individual cases of transaminase and bilirubin elevations and an overall increased rate of transaminase elevation to some defined extent, e.g., a 2% rate of 3x ULN elevation.

3.   Information to Be Collected For Each Drug on Lists #1 and #2.

To evaluate the signals, for drugs on List #1 or #2, the pre-marketing data base should be examined as described in Appendix I, which is essentially a list of what a reviewer of a new drug application (NDA) should consider during drug evaluation and review. In addition, there should be a brief summary of the metabolism of the drug. With these data, it should be possible to describe the performance characteristics of the currently identified potential signals (Hy's Law, serum transaminase elevations of various extents and frequencies of occurrence) and make reasonable estimates of the risk a given sample size could exclude.

Appendix I

Information to be Extracted from Pre-Marketing Databases

I.   Data Collection

A.   Overview of liver test data (tests performed, frequency, specific follow-up plans if abnormalities seen).

B.   Specific follow-up plan if test is elevated at end of treatment.

C.   Re-challenge plan, if any.

D.   Exclusions from studies because of liver test abnormalities, if any.

II.   Observations and Analyses

A.   Abnormalities in controlled trials (separate for pooled placebo controlled, active controlled) with greater than two week exposure. Rates of occurrence may be given as events/exposed; positive findings can be also analyzed as events per patient year and examined for rates over time.

1.   Rates of >3x ULN, 5x, 10x or 20x elevation of ALT, AST, and either AST or ALT.

2.   Rates of any elevations of bilirubin, rate of elevated bilirubin to >1.5x ULN, to >2x ULN.

3.   Rates of ALP >1.5 x ULN

All rates should be given for both drug and control group. Normal ranges for all tests should be provided.

B.   Total database of persons with exposure > two weeks.
Same as for controlled database.

C. Individual Events

1. Listing of patients with any elevated transaminase (>3x ULN) (without marked ALP elevation) associated with increase of bilirubin to >ULN.

2. Show time course of enzyme and bilirubin elevations.

3. For each, review clinical situation
   a. Ethanol history;
   b. Evidence of viral hepatitis, or other known liver disease;
   c. Symptoms and course;
   d. Special studies, notably liver biopsy results;
   e. Possible confounding, including history of liver disease, concomitant medications (acetaminophen, other hepatotoxins, herbal products); and
   f. Obesity (weight, height, body mass index) or type II diabetes mellitus because of transaminase elevations caused by NASH (non-alcoholic steatohepatitis).