Some points from the July 2009 “Guidance to Industry: Drug-Induced Liver Injury, Premarketing Clinical Evaluation” for consideration and discussion at the evening session 20 March 2013:

- many favorable comments have been received about this guidance, and some feel it needs no revision or change in 2013;

- nevertheless, it was written by a team that assumed normal, healthy livers in subjects being studied in clinical trials before approval of the drug for marketing, which will NOT be the case in considering the many submissions for new drugs and combinations for treating and possibly almost curing chronic hepatitis C, treatments of other viral hepatitides, oncology patients who may have liver metastases, and other conditions where the liver is not normal;

- there is a heavy reliance on the height of elevation of serum enzyme activities such as ALT as an indicator of the severity of the liver injury, following the probably erroneous assumption of this by the consultants advising the National Cancer Institute in their initial formulation of the Critical Toxicity Criteria for Adverse Events in 1982, and in subsequent versions up to 2009 (version 4). It is far more likely that the true assessment of severity of liver injury should be based on whole organ dysfunction (as measured by the total bilirubin concentration in serum, BILT, or the plasma prothrombin time or its derivative international normalized ratio, INR, that are the only two true measures of whole organ dysfunction that are commonly done. Serum enzyme activities are misnamed as liver function tests but do not really measure any liver function whatsoever, only the rate and extent of hepatocellular injury and should be assessed as indicators of urgency to recheck and follow the values rather than as function tests;

- the present guidance does not give much attention to the well known fact that the liver is a very robust organ that can withstand and recover from quite severe injury, such as 65% resection, and even without showing serious dysfunction can change itself, adapt, become tolerant to doses of a drug that may have caused some initial injury, as seen frequently with drugs such as isoniazid. Only the rare patient, 1 or 2 per 1000, lacks the ability to accomplish such adaptation and must have the drug stopped before very serious, irreversible injury and liver failure occur (this is mentioned a bit in section A8, on rechallenge);

- there is no discussion of the value of narrative summaries for making the diagnosis of most likely cause of the findings observed, nor of the value of inspecting the time course of all liver tests done in a given patient suspected of DILI by initial finding of elevated peak values of \{ALT>3xULN & BILT>2xULN\}, leading to misguided attempts by statisticians to “diagnose” DILI based simply of peak elevations of those two chemical measures.

- there is no discussion of what constitutes a valid baseline, one point not being sufficient to establish a line of any sort, nor of what constitutes ”normal” values for tests.

Proposed “Best Practices” and a Liver Safety Research Consortium,
Points raised at the 9 November 2013 meeting preceding the annual meeting of the American Association for the Study of Liver Diseases (AASLD) called by Drs. Michael Merz, Mark Avigan, and Paul Watkins, representing industry, government, and academia, as a Liver Safety Research Consortium (LSRC). There were 52 interested persons who came, listened, and debated points about developing some possible “Best Practices” for all three groups to observe in the design and evaluation of clinical trials, with respect to DILI (see list attached).

Some of the points discussed that may be worth further airing before the wider group that may assemble on the evening of 20 March 2013, and may be included in a paper to be written and published by the organizers of the 9 November meeting, included:

- there is need for standardization of nomenclature and definitions, for methods to capture and report data, and for systematic training of investigators;

- pre-study and baseline clinical and laboratory data, and storage of frozen specimens of sera or tissues, may enhance later retrospective studies to determine causes of problems;

- in treatment of chronic viral hepatitis (C and B, especially) the pre-treatment baseline values established may need to by modified to recognize a lower on-treatment baseline of values against which subsequent effects of treatment or other processes are compared;

- pooling of liver safety data across studies and protocols, may significantly enhance the understanding of analyses and allow new research insights;

- there is no specific biomarker or combination of tests useful for both preclinical and clinical diagnosis of DILI as probable (>50% likely, more than all other possible causes combined);

- narratives describing the clinical course of subjects or patients showing adverse liver effects should be written by physicians skilled in medical differential diagnosis for probable causality, ideally as close as possible to the place and time care and observation are given;

- terminology and nomenclature should be standardized and agreed upon in order to for a basis for establishing “Best Practices”;

- it would be advantageous to share data, pool similar data, in a “warehouse” that would allow research and take advantage of the unparalleled collection of clinical data submitted for review of potential new drugs, in the style pioneered by the cardiovascular folks.

In addition to the points above, the popular use of receiver operating characteristic (ROC) curves of sensitivity and specificity should be questioned, for tests done and “verified” on samples of subjects where prevalence or incidence of injuries are high, then applying the results for screening subjects and patients for rare events, in which the “predictive” (better, diagnostic, indicative) value of positive test results falls precipitously so that nearly all of the positive results are false unless tests of exquisitely high specificity are used. Better sensitivity is not enough!