ORIGINAL RESEARCH ARTICLE

Drug Saf 2012; 35 (10): 865-875
0114-5916/12/0010-0865/$49.95/0

Validation of Multivariate Outlier Detection Analyses Used to Identify Potential Drug-Induced Liver Injury in Clinical Trial Populations

Xiwu Lin,[1] *Daniel Parks*,[1] *Jeffery Painter*,[2] *Christine M. Hunt*,[3] *Heide A. Stirnadel-Farrant*,[4] *Jie Cheng*,[1] *Alan Menius*[2] and *Kwan Lee*[1]

1 Medical Analytics, GlaxoSmithKline, Collegeville, PA, USA
2 Medical Analytics, GlaxoSmithKline, Research Triangle Park, NC, USA
3 Global Clinical Safety and Pharmacovigilance, GlaxoSmithKline, Research Triangle Park, NC, USA
4 Worldwide Epidemiology, GlaxoSmithKline, Stockley Park, UK

Abstract

Background: Potential severe liver injury is identified in clinical trials by ALT >3 × upper limits of normal (ULN) and total bilirubin >2 × ULN, and termed 'Hy's Law' by the US FDA. However, there is limited evidence or validation of these thresholds in clinical trial populations. Using liver chemistry data from clinical trials, decision boundaries were built empirically with truncated robust multivariate outlier detection (TRMOD), in a statistically robust manner, and then compared with these fixed thresholds. Additionally, as the analysis of liver chemistry change from baseline has been recently suggested for the identification of liver signals, fold-baseline data was also assessed.

Objective: The aim of the study was to examine and validate the performance of fixed and empirically derived thresholds for severe liver injury in generally healthy clinical trial populations (i.e. populations without underlying renal, haematological or liver disease).

Methods: Using phase II–IV clinical trial data, ALT and total bilirubin data were analysed using outlier detection methods to compare with empirically derived and fixed thresholds of the FDA's Hy's Law limits, which were then assessed graphically with the FDA's evaluation of Drug-Induced Serious Hepatotoxicity (eDISH) assessing fold-ULN, as well as a modified eDISH (mDISH) to assess fold-baseline liver chemistries. Data from 28 phase II–IV clinical trials conducted by GlaxoSmithKline were aggregated and analysed by the TRMOD algorithm to create decision boundaries. The data consisted of 18 672 predominantly female subjects with a mean age of 44 years and without known liver disease.

Results: Among generally healthy clinical trial subjects, the empirically-derived TRMOD boundaries were approximately equivalent to 'Hy's Law'.

TRMOD boundaries for identifying outliers were an ALT limit of 3.4×ULN and a bilirubin limit of 2.1×ULN, compared with the FDA's 'Hy's Law' of 3×ULN and bilirubin 2×ULN. Inter-laboratory data variations were observed across the 28 studies, and were diminished by use of baseline-corrected data. By applying TRMOD to baseline-corrected data, these boundaries became ALT limit of 3.8×baseline and bilirubin limit of 4.8×baseline. Cumulative incidence plots of liver signals identified over time were examined. TRMOD analyses identified normative boundaries and outliers that provide comparative data to detect liver signals in similar trial populations.

Conclusions: TRMOD liver chemistry analyses of clinical trial data in generally healthy subjects have confirmed the FDA's Hy's Law threshold as a robust means of detecting liver safety outliers. TRMOD evaluation of liver chemistry data, by both fold-ULN and fold-baseline, provides complementary analyses and valuable normative data for comparison in similar patient populations. No liver signal is present when new clinical trial data from similar patient populations lies within these normative boundaries. Use of baseline-corrected data diminishes inter-laboratory variation and may be more sensitive to possible drug effects. We suggest examining liver chemistries using graphical depictions of both ULN-corrected data (eDISH) and baseline-corrected data (mDISH), as complementary methods.

Background

Drug-induced liver injury is the most frequent cause of postmarketing drug withdrawal,[1] commonly terminates promising drug development programmes and is the foremost cause of acute liver failure in the US.[2] To address these issues, the US FDA premarketing liver safety guidance provides subject safeguards through its standardized approach to liver safety and liver chemistry subject stopping criteria, where study drug treatment is stopped when pre-specified liver chemistry threshold criteria are reached.[1] Earlier and more sensitive detection of drug-induced liver injury and intervention during drug development is likely to enhance clinical safety.

Hy Zimmerman[3] reported that 10–50% of patients with drug-induced hepatocellular jaundice develop fatal liver failure. These data have been confirmed in large population studies reporting 9–12% mortality with drug-induced hepatocellular injury with jaundice.[4,5] The FDA-coined term 'Hy's Law'[1,6,7] is used to identify liver chemistry elevations of serious clinical concern: ALT >3× upper limits of normal (ULN) and total bilirubin >2×ULN. Furthermore, the FDA guidance cautions that "Hy's Law [is an] ominous indicator [of] a drug likely to cause severe drug-induced liver injury".[1] Recently, FDA researchers have developed a tool for evaluation of Drug-Induced Serious Hepatotoxicity (eDISH)[8] to identify and help predict drug-induced liver injury. eDISH is essentially a graphical depiction of Hy's Law, implementing its fixed decision boundaries. The eDISH graph effectively organizes the liver chemistry data by displaying peak serum ALT and bilirubin levels for each clinical trial subject. eDISH was recently applied to phase III rivaroxaban clinical trials.[9] Watkins et al.[9] concluded that eDISH is a highly efficient and effective way to examine and summarize serum ALT and bilirubin levels from randomized controlled clinical trials.

Because of the clinical importance of Hy's Law in drug development, we examined its performance in clinical trial populations. Earlier

analyses of phase II–IV clinical trial data from more than 18 000 predominantly female patients without known liver disease were employed;[10] females are disproportionally affected by severe drug-induced liver injury.[11] These analyses revealed that no subjects met Hy's Law criteria during treatment or follow-up when receiving investigational compounds without liver signals.[10] These analyses also examined liver chemistry elevations, using fold-upper limits of normal (or × ULN), and found the rates of liver chemistry elevations were similar in patients receiving active drug to those in placebo recipients.

The value of assessing change from baseline liver chemistry data during phase I clinical trials has recently been reported by Cai et al.[12] and M'Kada et al.[13] Baseline-corrected data assists in eliminating undesirable inter-laboratory variation, and hence may be more sensitive to the effects of underlying drug effects.[14] Baseline-corrected data is frequently used in drug efficacy analyses. However, unlike the ULN-corrected data, the decision boundaries based on the baseline-corrected (fold-baseline) data of ALT and bilirubin have not been established.

Robust multivariate analyses of liver chemistry data have been examined for early identification of drug-induced liver injury.[15] Recently, Lin et al.[16] proposed a truncated robust distance for assessing clinical laboratory safety biomarkers which is multivariate and robust to outliers, handling several biomarkers and their correlations simultaneously. Extremely small values of safety markers are generally clinically insignificant and truncation allows automatic exclusion of these clinically irrelevant outliers, thus focusing analyses on patients with clinically significant changes. This method applied to outlier detection of safety markers is known as Truncated Robust Multivariate Outlier Detection (TRMOD). Unlike the fixed boundaries of Hy's Law and eDISH, the statistical property of a TRMOD decision boundary can be obtained since TRMOD boundaries are flexible and determined by a given data and can be adjusted readily by controlling the corresponding tolerance probability of confidence.

Our study goals were to apply TRMOD to liver chemistries to assess and characterize the statistical properties of decision boundaries (tolerance limits) of Hy's Law and eDISH, examining both fold-ULN and fold-baseline data. To achieve these goals, we first applied TRMOD to liver chemistry data from an aggregated clinical trial database to obtain decision boundaries and thresholds for ALT and bilirubin and compare them with those of Hy's Law. We then applied TRMOD to baseline-corrected data from the same aggregated clinical trial database to obtain decision boundaries and thresholds for fold-baseline data of ALT and bilirubin. Based on the thresholds for fold-baseline ALT and bilirubin data, we proposed a modified eDISH (mDISH) as a complementary tool to eDISH for detecting liver signals in other similar trial populations.

Methods

Truncated Robust Multivariate Outlier Detection (TRMOD) for Liver Chemistries

Identification of abnormal liver chemistry data may be considered as an outlier detection problem. An outlier refers to an observation that deviates markedly from the pattern or distribution of the majority of the data, and is usually identified through robust statistical methods. Multivariate outlier detection based on a robust distance has been studied extensively[17] and applied to detect outliers in multivariate laboratory data.[15] Mahalanobis Distance[18] measures the distance of a subject from the centre of the data defined by correlated multivariate variables, assumed to be normally distributed. Robust distance is obtained by using the robust estimate of mean and covariance in the calculation of Mahalanobis Distance. In essence, the robust estimates assume that the majority of data come from a multivariate normal distribution, but that some outliers exist. Relaxing the assumption of normality in this way enables the outliers to be more easily identified. Subjects with robust distance greater than a given cutoff will be considered as outliers. The decision boundary for multivariate outlier detection based on multivariate normal distribution has an ellipsoidal shape in general[19] and is an ellipse for the bivariate (two markers) case (figure 1a). The

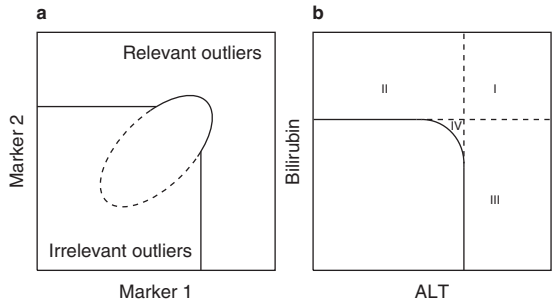

Fig. 1. TRMOD boundaries in (**a**) a typical scenario, and (**b**) for the four outlier regions when ALT and bilirubin are the biological markers: region I = severe toxicity; region II = elevated bilirubin; region III = elevated ALT; and region IV = potentially severe toxicity. **TRMOD** = truncated robust multivariate outlier detection.

cutoff of robust distance or the decision boundary can be adjusted based on a specified false detection probability value. The false detection probability here refers to the probability of the data generated from the underlying multivariate normal distribution (i.e. non-outlier subject) falling outside of the decision boundary. The 'slope' of the ellipse gets steeper as the correlation between the two measurement increases.

Multivariate outliers detected based on such decision boundaries will include outliers in all directions. However, only abnormally high elevations of liver chemistry measurements indicate a potential liver safety issue. Hence, the outliers with abnormally small values of liver chemistries are not of interest in identifying potential liver toxicity and would be considered clinically irrelevant outliers. TRMOD[16] was proposed as a robust statistical method for identification of clinically relevant outliers in laboratory safety data while automatically excluding clinically irrelevant outliers. The TRMOD method eliminates clinically irrelevant outliers in any variable by fixing their values at a truncation point near the centre of the data before computing their robust distances. Specifically, the value of the truncated point will be used in the calculation of a robust distance if a variable has a value less than the value of the truncated point for the corresponding variable. Subjects with truncated robust distance greater than a given cutoff will be considered as clinically relevant outliers. The truncation effectively blocks irrelevant outliers and the decision boundary defined by the truncated robust distance is shown in figure 1a as a solid line.[16] The TRMOD boundaries can be adjusted based on specified false detection probability value. A false detection probability of 0.001 (or 1 in 1000) means only about 0.1% of data randomly selected from the underlying normal distribution may be expected to be outside of the decision boundary. In other words, if all the data are randomly sampled from the normal distribution and hence no true outlier subject is presented, we would expect about 0.1% of the data to fall outside the decision boundary when a false detection probability of 0.001 is used. The choice of value for false detection probability will depend on an acceptable level defined by clinical practitioners. However, the false detection probability in most cases may be limited to very small values such as 1 in 1000 or 1 in 10 000. In applying TRMOD to liver chemistry data, log transformation of ALT and bilirubin will be used so that the majority of the data can be modelled approximately as a multivariate normal distribution. The two liver chemistry measurements, ALT and bilirubin, are not highly correlated, and the decision boundary then appears rounded, similar to figure 1b.

An important feature of the TRMOD boundaries involves the truncation lines, which are determined by their intersection with the horizontal and vertical axes. We will call them the x-limit and y-limit for future referencing, where x and y indicate coordinates of the horizontal and vertical axes, respectively. A very important feature of the TRMOD boundary information is represented by the (x-limit, y-limit), which can be used for comparison with other thresholds based on methods such as eDISH. By extending the truncation line as given in figure 1b, it is possible to achieve decision boundaries very similar to eDISH using TRMOD. For the two liver chemistry measurements, ALT limit will be interpreted as the x-limit and the bilirubin limit as the y-limit. ALT and bilirubin data can be either ULN-corrected or baseline-corrected. ALT and bilirubin limits define regions similar to eDISH. Together, they form regions identified as Region I, severe toxicity or potential Hy's Law; Region II, elevated bilirubin; Region III, elevated ALT; and

Region IV, potential toxicity. Outliers lying in region IV as a result of the multivariate analysis may indicate some abnormality in both ALT and bilirubin simultaneously which may require further attention. In fact, if we ignore region IV, the shape of the decision boundaries are exactly the same in both TRMOD and eDISH. An important difference, however, is that the TRMOD boundaries are estimated from data (data for our analyses were from generally healthy patients enrolled in clinical trials), compared with the eDISH boundaries which are fixed since they were derived from Hy's Law.

TRMOD decision boundaries and thresholds were computed for two liver safety biomarkers, ALT and bilirubin, for a generally healthy patient population participating in clinical trials,[10] applying both fold-ULN and fold-baseline. The 95% confidence intervals (CIs) of ALT limit and bilirubin limit were estimated through use of a repeated bootstrapping method (random sampling with replacement). Once the TRMOD boundaries and thresholds are computed, they were compared with those proposed by Hy's Law and eDISH fixed boundaries, and then comparisons of statistical properties were performed. Additionally, the TRMOD boundaries based on fold-baseline data were used to form an mDISH method.

Generally Healthy Patient Populations

In order to estimate the decision boundaries of TRMOD, liver chemistry data of 18 672 generally healthy patients from 28 phase II–IV trials conducted by GlaxoSmithKline were used.[10] The details on how these trials are selected are described in reference.[10] Generally healthy patients refer to patients without underlying renal, haematological or liver disease. Of these patients, 92.3% of patients were female, with the male patients (n = 1400) having graded ALT elevations during treatment and follow-up that were similar to those of the female participants. The mean age of patients was 44.3 years. These patients are representative of many therapeutic area populations (e.g. migraine, osteoporosis, functional bowel disease, etc.).

Results

Hy's Law, Evaluation of Drug-Induced Serious Hepatotoxicity (eDISH) Decision Boundary for Liver Safety Data

The peak ULN-corrected ALT and bilirubin levels of 18 672 generally healthy patients are plotted on eDISH (figure 2). Potential severe liver injury can be identified by ALT >3 × ULN and bilirubin >2 × ULN, and is depicted graphically as the potential Hy's Law region in the upper right quadrant.[8] Isolated bilirubin elevations are generally due to an innocuous bilirubin conjugating defect, termed Gilbert's syndrome,[20] and appear in the 'Gilbert's cholestasis' quadrant. The 'Temple's corollary' region displays ALT >3 × ULN, observed frequently in drugs associated with hepatotoxicity. And finally, the 'Normal' region is occupied by most subjects without drug-associated liver injury. No patient data are found in the potential Hy's Law region.

TRMOD Boundaries Estimated from Generally Healthy Patients' Data

The TRMOD and eDISH decision boundaries both look alike and are comparable. Therefore, the TRMOD boundaries were estimated based on data from the generally healthy patient populations,

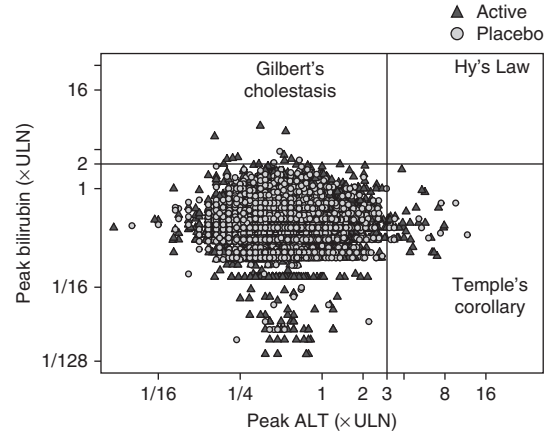

Fig. 2. eDISH plot: scatter plot of peak ULN-corrected ALT and bilirubin with eDISH decision boundaries showing region I, Hy's Law; region II, Gilbert's cholestasis; and region III, Temple's corollary. **eDISH** = evaluation of Drug-Induced Serious Hepatotoxicity; **ULN** = upper limit of normal.

Table I. Upper limit of normal-corrected ALT and bilirubin limits with 95% CIs based on 500 bootstrapping samples

False detection probability	ALT limit (95% CI)	Bilirubin limit (95% CI)
1 in 1000	2.5 (2.4, 2.6)	1.55 (1.5,1.6)
1 in 10 000	3.4 (3.1, 3.6)	2.1 (1.9, 2.3)

with the main focus being the ULN-corrected ALT limit and bilirubin limit. Note that for potential 'Hy's Law', the ULN-corrected ALT and bilirubin limits are 3 and 2, respectively. The TRMOD ULN-corrected ALT and bilirubin limits with a 95% CI based on 500 bootstrapped samples for two different false detection probabilities of 1/1000 and 1/10 000 are summarized in table I. From table I, the TRMOD ULN-corrected ALT and bilirubin limits of 3.4 and 2.1, respectively, were slightly higher than eDISH limits 3 and 2, respectively, which made the TRMOD boundary for false detection probability of 1/10 000 slightly wider than eDISH (figure 3). No patient data appear in region I (potential Hy's Law quadrant).

Furthermore, in table I, it is apparent that Hy's Law limits lie somewhere between the estimated TRMOD limits calculated with false detection probabilities of 1/1000 and 1/10 000, but was closer to limits estimated with a false detection probability of 1/10 000 on average. Hence, eDISH, which is based on Hy's Law, performs similarly to TRMOD in its functionality.

Outlier Analysis Based on Baseline-Corrected Data

Baseline-corrected data (fold-baseline) was also examined and compared with ULN-corrected data (fold-ULN). Raw data, ULN-corrected data and baseline-corrected data of ALT by different ULN values are plotted in figure 4. Differing ULN values approximately correspond to the varying laboratories performing the ALT measurements, which revealed numerous inconsistencies across the laboratories. Both the raw and ULN-corrected ALT show high variability among the laboratories. In contrast, baseline-corrected data demonstrate consistency across the laboratories and an average close to 0 (in log scale), meaning ALT values stay the same on average as the baseline during the trial. Similar performance was noted for ULN-corrected and baseline-corrected bilirubin values. In summary, baseline-corrected data is more consistent across laboratories compared with ULN-corrected data. Therefore, baseline-corrected data would be more sensitive to drug effects and are complementary to ULN-corrected data.

Next, the TRMOD was applied to baseline-corrected data to estimate the decision boundary, and compared with the eDISH boundary of (3, 2) using a false detection probability of 1 in 10 000. The TRMOD limit for baseline-corrected ALT was estimated as 3.8 (95% CI 3.5, 4.0) and the limit for baseline-corrected bilirubin was 4.8 (95% CI 4.4, 5.2). The baseline-corrected ALT limit of 3.8 and bilirubin limit of 4.8 are used in our mDISH, which applies baseline-corrected data. Figure 5 shows the mDISH plot of peak baseline-corrected ALT and bilirubin data. Based on figure 5, data of one patient appears in the potential Hy's Law quadrant (region I), suggesting potentially serious liver chemistry elevations from baseline.

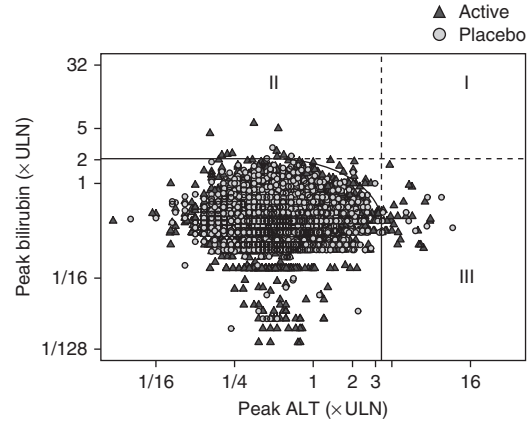

Fig. 3. Scatter plot of peak ULN-corrected ALT and bilirubin with estimated TRMOD boundaries based on ULN-corrected data from generally healthy patients participating in clinical trials given false detection probability of 1/10 000 showing region I, severe toxicity (type I); region II, elevated bilirubin (type II); region III, elevated ALT (type III). **TRMOD** = truncated robust multivariate outlier detection; **ULN** = upper limit of normal.

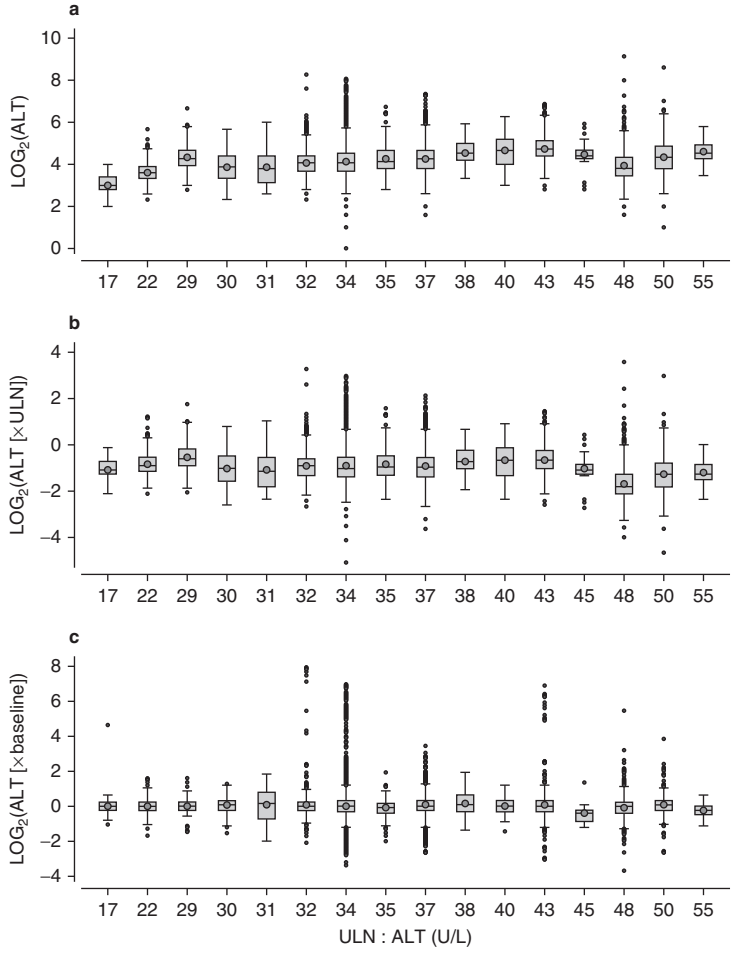

Fig. 4. Box plots of (**a**) raw ALT (in log transformation) by each ULN value; (**b**) ULN-corrected ALT (in log transformation) by each ULN value; and (**c**) baseline-corrected ALT (in log transformation) by each ULN value. The box within a box plot shows the range from the lower to upper quartile, with the grey circle and the solid line inside the box indicating the mean and median, respectively. **ULN** = upper limit of normal.

Modified eDISH as a Complementary Tool to eDISH

We can further compare the occurrence of liver signals over time between treatments after patients with liver signals are identified. The cumulative incidence plots display the appearance of liver signals in the three outlier regions of eDISH and mDISH over time in the generally healthy patient data (figure 6). This tool adds value by allowing the user to compare an experimental drug with a control or an historically 'liver safe' drug for assessing safety issues and monitoring liver chemistry during development. While liver chemistry elevations may occur at similar times in the active or placebo groups, as shown in figure 6, the delayed appearance of marked liver chemistry elevations in the treated groups is ominous and has been observed in drugs withdrawn from the market (e.g. troglitazone).[21] Analysis of clinical safety laboratory data using eDISH and mDISH, together with the cumulative incidence plots, could be used to enhance future regulatory submissions. A proportionally

similar, small number of subjects receiving active drug and placebo exhibited ALT ≥3×ULN (up to 12 months) in the Temple's corollary region (figure 6c). The 12-month cumulative incidence rates of ALT elevations from baseline are also similar for patients receiving active drug compared with those receiving placebo (figure 6f). Results such as these might be used to guide further investigation of a potential liver injury during a clinical trial.

Discussion

Using the TRMOD in generally liver-healthy clinical trial subjects, Hy's Law (ALT >3×ULN and bilirubin >2×ULN) and eDISH were shown to be simple and reasonable tools to detect potential liver safety outliers. The characteristics of Hy's Law limits and eDISH for detecting potential liver safety outliers in differing patient populations deserve further study. In this particular patient population, Hy's Law and eDISH limits are approximately equivalent to TRMOD boundaries when applied to fold-ULN data with mean ALT 3.4×ULN and bilirubin 2.1×ULN. Additionally, TRMOD methods allow us to produce normative liver chemistry data boundaries which can be used to compare new clinical trial data from similar patient populations, and also support the use of fold-baseline changes in the assessment of liver signals.

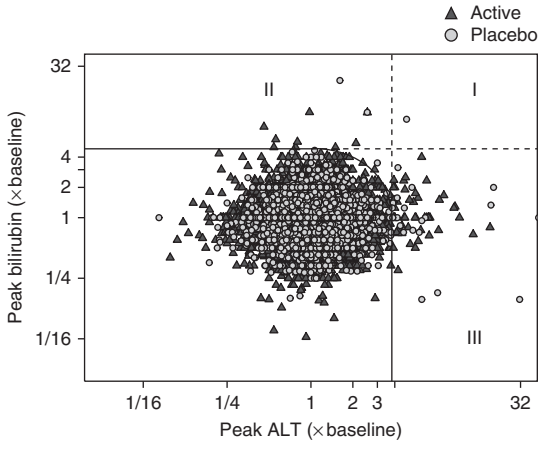

Fig. 5. mDISH plot: scatter plot of peak baseline-corrected ALT and bilirubin with estimated TRMOD boundaries based on baseline-corrected data from generally healthy patients participating in GlaxoSmithKline clinical trials given false detection probability of 1/10 000 showing region I, severe toxicity (type I); region II, elevated bilirubin (type II); region III, elevated ALT (type III). **mDISH** = modified evaluation of Drug-Induced Serious Hepatotoxicity; **TRMOD** = truncated robust multivariate outlier detection.

The flexibility of TRMOD allows it to be applied to baseline-corrected data as well as ULN-corrected data. With increasing interest in evaluating the change from baseline to minimize laboratory variation,[12,13] the mDISH method was developed by examining fold-baseline elevations. This baseline-corrected data may be more sensitive to possible drug effects. However, baseline-corrected data may be less sensitive than ULN-corrected data in populations with baseline liver chemistry elevations such as chronic hepatitis or non-alcoholic fatty liver disease.

mDISH is a graphical tool complementary to eDISH where both baseline-corrected peak ALT and baseline-corrected peak bilirubin of each patient are plotted. mDISH yielded a modified Hy's Law limit of ALT 4×baseline and bilirubin 5× baseline (compared with the FDA's Hy's Law of ALT >3×ULN and bilirubin >2×ULN). mDISH limits of ALT >4×baseline and bilirubin >5×baseline may be used for populations similar to the generally healthy patient populations used in this study. The greater analytic and biological variability of bilirubin due to Gilbert's syndrome,[22] and possible drug-related UDP-glucuronosyltransferase (UGT) inhibition,[23] likely accounts for the 2-fold greater variation in bilirubin observed in mDISH versus eDISH. mDISH is reliant on the baseline values. The use of a single baseline value for comparison introduces potential issues with analytic and biological variation,[22] suggesting the use of averaged pre-treatment baseline and screening values when available. While use of fold-baseline data appears promising, additional data is needed to confirm the value it adds to the analysis for detecting potential liver injuries.

The TRMOD approach, with use of both methodologies (eDISH and mDISH) to develop normative data and identify outliers, is likely to further enhance the detection of drug-induced liver injury and liver signals during drug development. For potential Hy's Law cases, trace plots can provide detailed case data on adaptation and

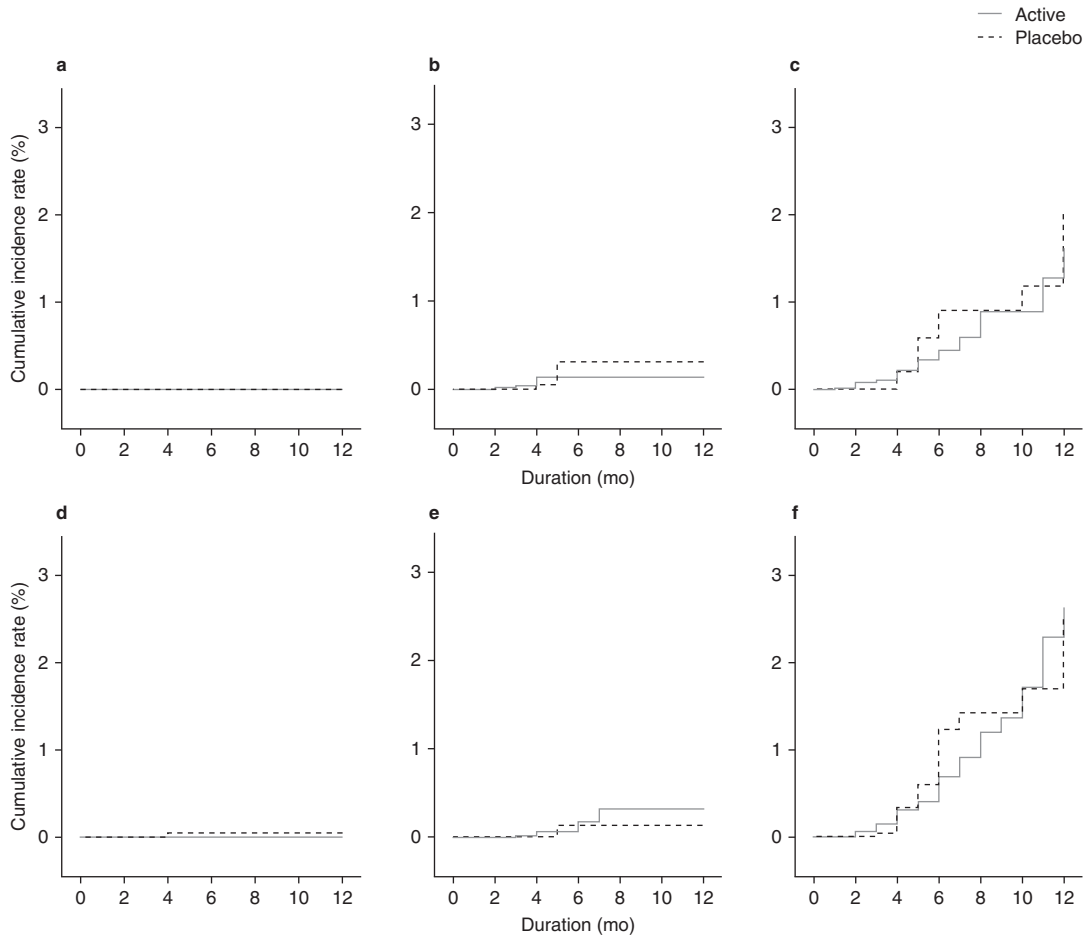

Fig. 6. Plots of cumulative incidence rates of potential liver injury signals identified by eDISH [(**a**) Hy's Law; (**b**) Gilbert's cholestasis; and (**c**) Temple's corollary] and mDISH [(**d**) type I; (**e**) type II; and (**f**) type III]. **eDISH** = evaluation of Drug-Induced Serious Hepatotoxicity; **mDISH** = modified evaluation of Drug-Induced Serious Hepatotoxicity.

liver chemistry course while additional clinical and laboratory data are obtained for case adjudication.[24] The combined use of the complementary eDISH (which assesses fold-ULN data) and mDISH (analysing fold-baseline-corrected data) methodologies is suggested in the evaluation of new clinical trial data.

The TRMOD approach can be used to establish a reference boundary to monitor patients during ongoing trials, identify outliers for detailed clinical evaluation and assess the appearance of liver signals during drug development. By establishing normative data in specific patient populations using both fold-ULN and fold-baseline, TRMOD also provides a valuable measure for comparing new clinical trial data from similar patient populations. However, the clinical utility of the TRMOD boundaries may need further validation since the boundaries are determined statistically.

The real utility of the multivariate outlier analysis tool (TRMOD) lies in its flexibility of application. It can be applied to any number of safety biomarkers, including cardiotoxicity, hepatotoxicity, nephrotoxicity and hematotoxicity. In this study, TRMOD was applied to aggregated liver

chemistries in generally healthy patient populations in clinical trials. Similarly, the methodology could be applied to patients with underlying liver disease or in those in various therapeutic areas (e.g. oncology patients). This will be the focus of future research efforts in applying TRMOD and mDISH.

The combined use of mDISH and eDISH may more confidently confirm normal liver chemistry data in similar patient populations. It may also, more sensitively, detect potential drug-induced liver injury cases. mDISH is likely to be of particular value for patient populations with underlying liver disease or baseline liver chemistry elevations, whereas TRMOD may be applied to baseline-corrected data. There are many unanswered questions remaining regarding the application of TRMOD to various therapeutic areas and patient populations. We look forward to continually expanding TRMOD's use and explore its role in drug safety.

Conclusions

TRMOD analysis of liver chemistry data in generally healthy clinical trial populations produced normative boundaries with limits similar to the FDA's Hy's Law threshold. Baseline-corrected data diminished inter-laboratory variation and may be more sensitive to possible drug effects than raw or ULN-corrected data. By applying TRMOD analysis to both fold-ULN and fold-baseline liver chemistry data, we can obtain valuable normative data for comparison in similar patient populations.

Acknowledgements

No sources of funding were used to conduct this study or prepare this manuscript. All authors of this manuscript are full time employees of GlaxoSmithKline.

References

1. US Department of Health and Human Services, FDA. Guidance for industry. Drug-induced liver injury: premarketing clinical evaluation [online]. Available from URL: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM174090.pdf [Accessed 2011 Apr 24]
2. Ostapowicz G, Fontana RJ, Schiodt FV, et al. Results of a prospective study of acute liver failure at 17 tertiary care centers in the United States. Ann Intern Med 2002; 137 (12): 947-54
3. Zimmerman HJ. Hepatotoxicity. 2nd ed. Philadelphia (PA): Lippincott Williams & Wilkins, 1999
4. Andrade RJ, Lucena MI, Fernández MC, et al. Drug-induced liver injury: an analysis of 461 incidences submitted to the Spanish Registry over a 10-year period. Gastroenterology 2005; 129: 512-21
5. Bjornsson E, Olsson R. Outcome and prognostic markers in severe drug-induced liver disease. Hepatology 2005; 42: 481-9
6. Senior JR. How can Hy's Law help the clinician? Pharmacoepidemiol Drug Saf 2006; 15: 235-9
7. Temple R. Hy's law: predicting serious hepatotoxicity. Pharmacoepidemiol Drug Saf 2006; 15: 241-3
8. Guo T, Gelperin K, Senior J. A tool to help you decide (detect potentially serious liver injury). March 2008 [online]. Available from URL: http://www.fda.gov/downloads/Drugs/ScienceResearch/ ResearchAreas/ucm076777.pdf [Accessed 2009 Dec 17]
9. Watkins PB, Desai M, Berkowitz SD, 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 2011; 34 (3): 243-52
10. Weil JG, Bains C, Linke A, et al. Background incidence of liver chemistry abnormalities in a clinical trial population without underlying liver disease. Regul Toxicol Pharmacol 2008; 52: 85-8
11. Reuben A, Koch DG, Lee WM, and the Acute Liver Failure Study Group. Drug-induced acute liver failure: results of a U.S. multicenter, prospective Study. Hepatology 2010; 52 (6): 2065-76
12. Cai Z, Christianson AM, Ståhle L, et al. Reexamining transaminase elevation in phase I clinical trials: the importance of baseline and change from baseline. Eur J Clin Pharmacol 2009; 65 (10): 1025-35
13. M'Kada H, Munteanu M, Perazzo H, et al. What are the best reference values for a normal serum alanine transaminase activity (ALT)? Impact on the presumed prevalence of drug induced liver injury (DILI). Regul Toxicol Pharmacol 2011; 60 (3): 290-5
14. Wooding WM. Planning pharmaceutical clinical trials: basic statistical principles. New York: Wiley-Interscience Publications, 1994
15. Southworth H. Detecting outliers in multivariate laboratory data. J Biopharm Stat 2008; 18: 1178-83
16. Lin X, Parks D, Zhu L, et al. Truncated robust distance for clinical laboratory safety data monitoring and assessment. J Biopharm Stat. In press
17. Rousseeuw PJ, Leroy AM. Robust regression and outlier detection. New York: Wiley, 1987
18. Maesschalck RD, Jouan-Rimbaud D, Massart DL. The Mahalanobis distance. Chemom Intell Lab Syst 2000; 50: 1-18
19. Trost DC. Multivariate probability-based detection of drug-induced hepatic signals. Toxicol Review 2006; 25: 37-54
20. Bosma PJ, Chowdhury JR, Bakker C, et al. The genetic basis of the UDP-glucuronosyltransferase 1 in Gilbert's syndrome. N Engl J Med 1995; 333: 1171-5

21. Graham DJ, Green L, Senior JR, et al. Troglitazone-induced liver failure: a case study. Am J Med 2003; 114: 299-306
22. Dufour DR, Lott JA, Nolte FS, et al. Diagnosis and monitoring of hepatic injury: I. Performance characteristics of laboratory tests. Clin Chem 2000; 46: 2027-49
23. Zhang D, Chando TJ, Everett DW, et al. In vitro inhibition of UDP glucuronosyl transferases by atazanavir and other HIV protease inhibitors and the relationship of this property to in vivo bilirubin glucuronidation. Drug Metab Dispos 2005; 33: 1729-39
24. Hunt CM, Papay JI, Theodore D, et al. Monitoring liver safety in drug development: the GSK experience. Regul Toxicol Pharmacol 2007; 49: 90-100

Correspondence: Dr *Xiwu Lin*, Medical Analytics, GlaxoSmithKline, 1250 South Collegeville Rd, Collegeville, PA 19426, USA.
E-mail: xiwu.2.lin@gsk.com