Setting Clinical Exposure Levels of Concern for Drug-Induced Liver Injury (DILI) Using Mechanistic in vitro Assays

Falgun Shah,*1 Louis Leung,† Hugh A. Barton,† Yvonne Will,† A. David Rodrigues,† Nigel Greene,*2 and Michael D. Aleo‡

*Worldwide Medicinal Chemistry, †Pharmacokinetics, Dynamics and Metabolism and ‡Drug Safety Research and Development, Pfizer Inc., Groton, Connecticut 06340

ABSTRACT

Severe drug-induced liver injury (DILI) remains a major safety issue due to its frequency of occurrence, idiosyncratic nature, poor prognosis, and diverse underlying mechanisms. Numerous experimental approaches have been published to improve human DILI prediction with modest success. A retrospective analysis of 125 drugs (70 = most-DILI, 55 = no-DILI) from the Food and Drug Administration Liver Toxicity Knowledge Base was used to investigate DILI prediction based on consideration of human exposure alone or in combination with mechanistic assays of hepatotoxic liabilities (cytotoxicity, bile salt export pump inhibition, or mitochondrial inhibition/uncoupling). Using this dataset, human plasma \( C_{\text{max,total}} \geq 1.1 \mu M \) alone distinguished most-DILI from no-DILI compounds with high sensitivity/specificity (80/73%). Accounting for human exposure improved the sensitivity/specificity for each assay and helped to derive predictive safety margins. Compounds with plasma \( C_{\text{max,total}} \geq 1.1 \mu M \) and triple liabilities had significantly higher odds ratio for DILI than those with single/dual liabilities. Using this approach, a subset of recent pharmaceuticals with evidence of liver injury during clinical development was recognized as potential hepatotoxicants. In summary, plasma \( C_{\text{max,total}} \geq 1.1 \mu M \) along with multiple mechanistic liabilities is a major driver for predictions of human DILI potential. In applying this approach during drug development the challenge will be generating accurate estimates of plasma \( C_{\text{max,total}} \) at efficacious doses in advance of generating true exposure data from clinical studies. In the meantime, drug candidates with multiple hepatotoxic liabilities should be deprioritized, since they have the highest likelihood of causing DILI in case their efficacious plasma \( C_{\text{max,total}} \) in humans is higher than anticipated.

Key words: drug-induced liver injury; plasma exposure; Liver Toxicity Knowledge Base; safety margin
idiosyncratic DILI is rare, not detected well in current preclinical toxicity settings, has obscure dose–response relationships, and can occur in individuals with genetic and/or environmental susceptibility factors (Roth and Ganey, 2010). The lack of knowledge of the susceptible population and contributing risk factors creates a challenge in predicting adverse reactions or serious outcomes in individual patients.

A number of in vitro high throughput screening assays have been developed to explore possible mechanisms contributing to DILI, for example, inhibition of mitochondrial function (Forcedu et al., 2012), hepatobiliary transporter inhibition (Dawson et al., 2012; Morgan et al., 2013), reactive metabolite formation and covalent binding (Nakayama et al., 2005; Usui et al., 2009), modulation of nuclear hormone receptor function (Wang et al., 2014), and cellular health (Greer et al., 2009; Xu et al., 2008). Some studies have shown correlation of total administered dose alone (Lammert et al., 2008) or in combination with drug lipophilicity (Chen et al., 2013a) with higher risk of DILI while others have examined combinations of mechanistic assays to better predict hepatotoxicity potential (Aleo et al., 2014; Thompson et al., 2012). Since liver injury has been reported with a large number of drugs, efforts have been undertaken to compile human hepatotoxicity data, including the National Institute of Health LiverTox Database (Hoofnagle et al., 2013) and the FDA Liver Toxicity Knowledge Base (LTKB) (Chen et al., 2011). These publicly available datasets have enabled development of new structure activity relationships for hepatotoxicity endpoints or triggered the development of knowledge-based and quantitative structure activity relationships (QSAR) models (Chen et al., 2014).

A majority of the aforementioned studies have not taken systemic exposure of a drug into consideration when predicting DILI. Since the liver is highly perfused and serves as the primary site for metabolism and clearance of xenobiotics, high drug exposure therein (as parent and/or metabolites) can potentially interfere with its function and increase the risk for injury. The objective of our study was to evaluate the role of drug exposure (plasma total or free drug concentrations) alone and in combination with mechanistic assays of hepatotoxicity liabilities in improving DILI prediction. In this study, we sought to address a number of questions using the annotated data as found in the FDA LTKB: (1) Is there a cut-off for plasma exposure (total or free) above which the incidence of severe DILI is significantly increased? (2) How does activity in mechanistic assays, such as cytotoxicity in transformed human liver epithelial (THLE) and human hepatocellular carcinoma (HepG2) cells, bile salt export pump (BSEP) inhibition and/or mitochondrial inhibition/uncoupling increase the predictivity for DILI risk when combined with plasma exposure? (3) What are the safety margins of concerns in these assays below which odds for severe DILI increase significantly? (4) Do pharmaceutical compounds that have failed in clinical trials, because of liver transaminase elevations, share similar characteristics such as higher plasma exposure and lower safety margins against one or more of the above mechanistic assays?

**MATERIALS AND METHODS**

The compounds used in this analysis were acquired from Sigma-Aldrich (St. Louis, Missouri), Toronto Research Chemicals (Toronto, Ontario), Selleck Chemicals (Houston, Texas), Sequoia Research Products (Pangbourne, UK), or from the Pfizer chemical bank (Groton, Connecticut).

Dataset. Information from the LTKB was retrieved from the US FDA website (Chen et al., 2011). A total of 287 drugs are classified into 3 categories: most-DILI ($n = 137$), less-DILI ($n = 85$), and no-DILI-concern ($n = 65$) using the DILI severity description in the FDA-approved prescription drug labels. Specifically, a drug was classified as most-DILI concern if it had been withdrawn/discontinued from the market or had a black box warning due to hepatotoxicity in the clinic or whose warning and precaution section carry a severe DILI description (eg, fatal hepatotoxicity, acute liver failure, liver necrosis, jaundice, hyperbilirubinemia, etc.). No-DILI drugs were those whose labels did not have any precautions for DILI and have been on the market for more than 10 years ensuring reliability in their labeling. Remaining drugs that had adverse reaction related to DILI but do not fall in either most- or no-DILI concerns were classified as “less-DILI” concerns (Chen et al., 2013a, b). Less-DILI compounds were not included in this analysis to determine the best separation between plasma exposure at risk for the 2 DILI extremes, no-DILI and most-DILI. Also the classification of less-DILI compounds is often debatable (Chen et al., 2013b). Additional no-DILI-concern compounds, defined using an identical classification system, were retrieved (Chen et al., 2013b). A total of 125 compounds were utilized in this analysis including 70 from the most-DILI-concern and 55 from the no-DILI-concern categories. The plasma $C_{\text{max}}$ total concentration ($C_{\text{max, total}}$) of a drug following single or multi-dose administration at a commonly prescribed dose or maximum prescribed daily dose were obtained from published literature (Xu et al., 2008). The majority of the plasma free fraction ($f_u$) data were collected from Obach et al. (2008) or determined internally. In addition, plasma protein binding data were obtained from Drug bank (http://www.drugbank.ca/), the FDA-online label repository (http://www.labels.fda.gov/), or published literature. Finally, plasma free $C_{\text{max}}$ concentrations ($C_{\text{max, u}}$) were derived by multiplying plasma $C_{\text{max, total}}$ with the corresponding free fraction ($f_u$) for a total of 107 compounds (64 most-DILI, 43 no-DILI).

**Measurement of cytotoxicity.** Cell viability was determined by measuring cellular ATP content following incubation of test compounds for 72 h in THLE and HepG2 cell lines (Greene et al., 2010; Marroquin et al., 2007). All compounds were tested in duplicate in THLE or HepG2 cell lines (<8% coefficient of variation, unpublished results). An 11-point concentration–response curve was generated over a range of 0.3–300 μM and data were reported in the form of inhibition concentration at which half maximal response ($IC_{50}$) was observed.

**Measurement of mitochondrial function.** Mitochondrial activity was assessed by measuring the oxygen consumption in isolated rat mitochondria to detect both the inhibition and uncoupling of oxidative phosphorylation using a phosphorescent oxygen-sensitive probe in a high-throughput fashion as previously described (Hynes et al., 2006). The highest concentration tested in mitochondrial assay was 25 μM for inhibition and 100 μM for mitochondrial uncoupling. The inter assay variability of the isolated mitochondrial oxygen consumption assay has been published elsewhere and assay results were found to be robust and reproducible (Hynes et al., 2013).

**Measurement of BSEP activity.** BSEP inhibition was conducted by measuring the transport of $[^{3}H]$taurocholic acid in human BSEP expressing vesicles (SB-BSEP-S9-VT or SB-BSEP-H15-VT) as previously described (Kostrubsky et al., 2006). Glyburide and spirinolactone were used as positive controls. The highest
concentration tested was 100 μM. The assay variability analysis on BSEP shows 90% of all pairs of replicates analyzed had variability <3-fold (unpublished results).

Data analysis. Pipeline Pilot (version 9.1, Accelrys, Inc., San Diego, California) custom scripting was used for data manipulation and analyses. The results were visualized using the TIBCO Spotfire 3.3 DXP. The IC50 in the cytotoxicity, mitochondrial, and BSEP inhibition assays was divided by the maximum plasma total drug concentration \( (C_{\text{max,total}}) \) to calculate the margin of safety for each assay for a given compound. We examined the contribution of mechanistic information to discriminate compounds classified as most-DILI-concern from those with no-DILI-concern using IC50 values in the mechanistic assays and safety margins at various thresholds. For the THLE and HepG2 cytotoxicity assays, the lowest IC50 for a compound in either assay was used. Both cell lines were valuable as we found they capture cytotoxicity potential of different chemotypes (acids vs bases) (Shah et al., 2014). Likewise for impairment of mitochondrial function, the minimum of either inhibition (IC50) or uncoupling activity (UC50) at a particular threshold was used. A truth table [true positives (TP), false positives (FP), true negatives (TN), false negatives (FN)] was generated for a binary classification of most-DILI- and no-DILI-concern compounds. Statistical analysis was carried out to calculate sensitivity (defined as the fraction of correctly predicted positives to all true positives for DILI, \( \text{TP}/(\text{TP}+\text{FN}) \)) or positive rate, specificity (defined as the fraction of correctly predicted negatives to all true negatives for DILI, \( \text{TN}/(\text{TN}+\text{FP}) \)) and false positive rate (1-specificity) at different thresholds. The receiver operating characteristics (ROC) curves were generated in Pipeline Pilot. The ROC curve measures the effectiveness of a potential predictor for a binary outcome. In this plot, the optimal models are found in the upper left region indicative of high sensitivity and high specificity and the diagonal represent the random model line. The ROC curve was created using the sensitivity and specificity for all cut-offs for exposure data (total and free) or selected thresholds for in vitro assays. In addition, other statistical parameters such as relative odds (TP/FP or FN/TN), odds ratio \((\text{TP}/\text{FP})/(\text{FN}/\text{TN})\), confidence interval for odds ratio calculated using online tool (https://www.medcalc.net/tests/odds_ratio.php), and area under the curve (AUC) for ROC curve were computed.

RESULTS

A complete dataset of 125 compounds with clinical DILI annotations, fraction unbound \( (f_u) \), plasma exposure \( (C_{\text{max,total}} \text{ or } C_{\text{max,u}}) \) are shown in Supplementary Table 1. First, we examined the relationships between plasma exposure \( C_{\text{max,total}} \) or \( C_{\text{max,u}} \) for differentiating most-DILI- from no-DILI-concern drugs. The ROC curve using sensitivity and specificity value over a continuum of cut-off values for plasma \( C_{\text{max,total}} \) and \( C_{\text{max,u}} \) were generated. Using the ROC plot (Fig. 1), we identified plasma \( C_{\text{max,total}} \geq 1.1 \mu M \) as having specificity of 73% and sensitivity of 80% for this set of compounds. The separation of most-DILI- and no-DILI-concern compounds \((n = 107 \text{ with } f_u \text{ data}) \) was less striking using \( C_{\text{max,u}} \) where an optimum cut-off for \( C_{\text{max,u}} = 0.51 \mu M \), specificity of 74% and only a modest sensitivity of 52% were achieved (Fig. 1, red line). In addition, the AUC was used to summarize the ROC curve as an alternative performance metric (Wager et al., 2013). In the best scenario, when a predictor perfectly separates 2 potential outcomes, the AUC would be 1 compared with 0.5 for a random predictor. The ROC AUC for plasma \( C_{\text{max,total}} \) model was found to be 0.78 and plasma \( C_{\text{max,u}} \) model was 0.66, respectively, emphasizing that plasma \( C_{\text{max,total}} \) is better at separating most-DILI versus no-DILI compounds than plasma \( C_{\text{max,u}} \) using the current dataset. The relative odds of most-DILI-concern compounds when the plasma \( C_{\text{max,total}} \) value is \( \geq 1.1 \mu M \) equals 3.73 compared with 0.47 when plasma \( C_{\text{max,u}} \) is \( <1.1 \mu M \). This means compounds with a plasma \( C_{\text{max,total}} \geq 1.1 \mu M \) are 10.7 times (odds ratio) more likely to represent most-DILI concern compounds in the clinic than
compounds from the current set with a plasma C\text{max,total} of <1.1\,\text{mM}. In contrast, compounds from the current dataset are only 3.1 times (odds ratio) more likely to be most-DILI if free exposure is >0.51\,\text{mM} than compounds below their respective values. Hence, only the C\text{max,total} was used in subsequent analyses since it was the best exposure parameter separating compounds between these 2 DILI classes.

We investigated if the ability to predict compounds classified as most-DILI could be improved by normalizing IC\text{50} in the mechanistic assays, such as cytotoxicity assessment, mitochondrial impairment, or BSEP inhibition, with plasma C\text{max,total}. Figure 2 demonstrates the separation of most-DILI (n = 38) from no-DILI-concern (n = 46) compounds without and with consideration of plasma C\text{max,total} using 4 different values for IC\text{50} (10, 25, 50, and 100\,\text{mM}) or safety margins (10-, 25-, 50-, 100-fold). Only 84 (n = 38 = most-DILI; 46 = no-DILI) compounds with plasma C\text{max,total} < 10\,\text{mM} were used in this analysis to enable a fair comparison as relatively few no-DILI compounds (n = 9 vs n = 32 most-DILI) from this dataset had plasma C\text{max,total} > 10\,\text{mM}. As depicted in the ROC curve, a model for the cytotoxicity assays without consideration of plasma C\text{max,total} (shown in orange, Fig. 2) is close to the random line with nearly identical true positive rate and false positive rate at all IC\text{50} values. On the contrary, when IC\text{50} in cytotoxicity cell lines are normalized with plasma C\text{max,total} (cytotoxicity margin, yellow line, Fig. 2), the true positive rate is improved with a significant reduction of the false positive rate at all assay exposure thresholds. Using a cytotoxicity threshold of 50, the sensitivity is 34% with a low false positive rate of 9% (specificity = 91%) suggesting a modest separation of most-DILI compounds from this dataset. Likewise, when IC\text{50} for BSEP or mitochondrial assays were normalized with plasma C\text{max,total}, the false positive rate (at all thresholds) and sensitivity (at thresholds < 100) were also improved. For example, at a threshold of 25 in the BSEP margin model (purple line, Fig. 2), the true positive rate is slightly increased from 21% (BSEP IC\text{50} value, cyan line in Fig. 2) to 29% with a false positive rate of 4%. For mitochondrial impairment (inhibition or uncoupling) at a threshold of 100 the FPR is reduced to 2% (vs 17% at 100\,\text{mM}) with a slight reduction in sensitivity at this threshold (29 vs 32%, pink and red lines for safety margin and IC\text{50} model for mitochondrial impairment, respectively). These data suggest that normalizing with plasma C\text{max,total} improves the true positive rate and false positive rate for most-DILI prediction using the assays analyzed.

Next we investigated why a subset of compounds with moderate to weak activity in these assays cause severe DILI whereas others have no evidence for DILI in clinic. Figures 3A–E shows the IC\text{50} of most- and no-DILI compounds in the cytotoxicity, mitochondrial and BSEP assays plotted against their plasma C\text{max,total} such that both intrinsic risk (only y-axis in log scale) and safety margin (both axis in log scale) in these assays can be understood. Figures 3A, 3C, and 3E investigate the DILI potential of compounds in a given quadrant of assay IC\text{50} versus plasma C\text{max,total} whereas Figures 3B, 3D, and 3E depict safety margins regardless of any defined IC\text{50} or exposure cut-off. Table 1 shows statistics on each assay and the combination of these assays. A number of most-DILI compounds with moderate to weak activity in these assays cause severe DILI whereas others have no evidence for DILI in clinic. Figures 3A–E shows the IC\text{50} of most- and no-DILI compounds in the cytotoxicity, mitochondrial and BSEP assays plotted against their plasma C\text{max,total} such that both intrinsic risk (only y-axis in log scale) and safety margin (both axis in log scale) in these assays can be understood. Figures 3A, 3C, and 3E investigate the DILI potential of compounds in a given quadrant of assay IC\text{50} versus plasma C\text{max,total} whereas Figures 3B, 3D, and 3E depict safety margins regardless of any defined IC\text{50} or exposure cut-off. Table 1 shows statistics on each assay and the combination of these assays. A number of most-DILI compounds with moderate to weak activities in these assays cause severe DILI whereas others have no evidence for DILI in clinic. Figures 3A–E shows the IC\text{50} of most- and no-DILI compounds in the cytotoxicity, mitochondrial and BSEP assays plotted against their plasma C\text{max,total} such that both intrinsic risk (only y-axis in log scale) and safety margin (both axis in log scale) in these assays can be understood. Figures 3A, 3C, and 3E investigate the DILI potential of compounds in a given quadrant of assay IC\text{50} versus plasma C\text{max,total} whereas Figures 3B, 3D, and 3E depict safety margins regardless of any defined IC\text{50} or exposure cut-off. Table 1 shows statistics on each assay and the combination of these assays. A number of most-DILI compounds with moderate to weak activities in these assays cause severe DILI whereas others have no evidence for DILI in clinic.
FIG. 3. Activity in in vitro assays versus plasma $C_{\text{MAX, TOTAL}}$ for most-DILI concern (red) and no-DILI concern (green) compounds. Minimum cytotoxicity values in THLE (shown as diamonds) or HEPG2 (shown as circles) cells versus plasma $C_{\text{MAX, TOTAL}}$ in terms of (A) intrinsic activity (vertical line at IC$_{50}$ of 100 $\mu$M and horizontal line at plasma $C_{\text{MAX, TOTAL}}$ of 1 $\mu$M) and (B) safety margins are shown. Inactive compounds in cytotoxicity assays are shown as triangles. BSEP inhibition versus plasma $C_{\text{MAX, TOTAL}}$ in terms of (C) intrinsic activity (vertical line at IC$_{50}$ of 30 $\mu$M and horizontal line at plasma $C_{\text{MAX, TOTAL}}$ of 1 $\mu$M) and (D) safety margins are shown. Minimum of mitochondrial inhibition (shown as circles) or uncoupling values (shown as diamonds) versus plasma $C_{\text{MAX, TOTAL}}$ in terms of (E) intrinsic activity (vertical line at IC$_{50}$ of 25 $\mu$M and horizontal line at plasma $C_{\text{MAX, TOTAL}}$ of 1 $\mu$M) and (F) safety margins are shown. Inactive compounds in mitochondrial assays are shown as triangles. Dotted lines in (B), (D), (F) show the region of plot with $<10$, $10$-$50$, $50$-$100$, or $>100$-fold margin. Full color version available online.
compound with a cytotoxicity IC$_{50}$ $<$100$\mu$M and plasma C$_{\text{max,total}}$ $>$1$\mu$M is 9.87 versus only 0.29 when cytotoxicity IC$_{50}$ $<$100$\mu$M and plasma C$_{\text{max,total}}$ is $<$1$\mu$M for the current data-set. Likewise, 15 most-DILI compounds with BSEP IC$_{50}$ values $<$30$\mu$M and plasma total exposure $>$1$\mu$M (safety margins $<$30-fold, Fig. 3C). Using this threshold, the odds ratio for most-DILI prediction was 31 versus 0.15 if plasma C$_{\text{max,total}}$ $<$1$\mu$M at BSEP IC$_{50}$ of $<$30$\mu$M (Table 1). A similar trend was observed for mitochondrial impairment assays whether inhibition or uncoupling was examined (Figs. 3D and 3E). At mitochondrial inhibition or
uncoupling IC$_{50}$ or UC$_{50} < 25 \mu M$ and plasma $C_{\text{max, total}} > 1 \mu M$, 13 most-DILI compounds were identified whereas none of the no-DILI compounds were present in this quadrant (odds ratio of 26 as opposed to 2.4 when plasma $C_{\text{max, total}} < 1 \mu M$).

Based on the present analysis and as shown in Table 1 (column labeled Safety Margin), statistically significant clinical exposure levels of concern for these mechanistic assays with higher odds ratio for most-DILI are $<30$-fold for BSEP, $<25$-fold for mitochondrial assays, and $<100$-fold for cytotoxicity assays; however, more conservatively, for BSEP and mitochondrial assays a $<100$-fold margin was also found to have higher odds ratio (ie, plasma exposure $>1 \mu M$ and IC$_{50} < 100 \mu M$, see Table 1). Also, as shown in Table 1, sensitivity for these assay/exposure combinations is between 19 and 33%, however, specificity in all
<table>
<thead>
<tr>
<th>Assay/liability</th>
<th>Threshold for Plasma C&lt;sub&gt;max,total&lt;/sub&gt; (µM) &amp; IC&lt;sub&gt;50&lt;/sub&gt; (µM)</th>
<th>Safety margin</th>
<th>Confusion matrix (TP:FP:FN:TN)</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Odds ratio</th>
<th>95% CI lower to upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSEP</td>
<td>&gt;1 µM &amp; &lt;30 µM</td>
<td>&lt;30-fold</td>
<td>15:055:55</td>
<td>0.21</td>
<td>1.00</td>
<td>31</td>
<td>1.81 to 530.9</td>
</tr>
<tr>
<td></td>
<td>&lt;1 µM &amp; &lt;30 µM</td>
<td>–</td>
<td>0:270:53</td>
<td>0.00</td>
<td>0.96</td>
<td>0.15</td>
<td>0.007 to 3.23</td>
</tr>
<tr>
<td></td>
<td>&gt;1 µM &amp; &lt;100 µM</td>
<td>&lt;100-fold</td>
<td>23:247:53</td>
<td>0.33</td>
<td>0.96</td>
<td>12.96</td>
<td>2.9 to 57.96</td>
</tr>
<tr>
<td></td>
<td>&lt;1 µM &amp; &lt;100 µM</td>
<td>–</td>
<td>1:769:48</td>
<td>0.01</td>
<td>0.87</td>
<td>0.09</td>
<td>0.011 to 0.83</td>
</tr>
<tr>
<td>Mitochondrial inhibition or uncoupling</td>
<td>&gt;1 µM &amp; &lt;25 µM</td>
<td>&lt;25-fold</td>
<td>13:057:55</td>
<td>0.19</td>
<td>1.00</td>
<td>26</td>
<td>1.51 to 449</td>
</tr>
<tr>
<td></td>
<td>&lt;1 µM &amp; &lt;25 µM</td>
<td>–</td>
<td>3:167:54</td>
<td>0.04</td>
<td>0.98</td>
<td>2.4</td>
<td>0.24 to 23.91</td>
</tr>
<tr>
<td></td>
<td>&gt;1 µM &amp; &lt;100 µM</td>
<td>&lt;100-fold</td>
<td>14:156:54</td>
<td>0.20</td>
<td>0.98</td>
<td>13.5</td>
<td>1.71 to 106.2</td>
</tr>
<tr>
<td></td>
<td>&lt;1 µM &amp; &lt;100 µM</td>
<td>–</td>
<td>3:767:48</td>
<td>0.04</td>
<td>0.87</td>
<td>0.3</td>
<td>0.07 to 1.24</td>
</tr>
<tr>
<td>Cytotoxicity in THLE or HEPG2 cells</td>
<td>&gt;1 µM &amp; &lt;50 µM</td>
<td>&lt;50-fold</td>
<td>8:162:54</td>
<td>0.11</td>
<td>0.98</td>
<td>6.96</td>
<td>0.84 to 57.50</td>
</tr>
<tr>
<td></td>
<td>&lt;1 µM &amp; &lt;50 µM</td>
<td>–</td>
<td>4:116:44</td>
<td>0.06</td>
<td>0.80</td>
<td>0.24</td>
<td>0.07 to 0.81</td>
</tr>
<tr>
<td></td>
<td>&gt;1 µM &amp; &lt;100 µM</td>
<td>&lt;100-fold</td>
<td>19:251:53</td>
<td>0.27</td>
<td>0.96</td>
<td>9.87</td>
<td>2.18 to 44.55</td>
</tr>
<tr>
<td></td>
<td>&lt;1 µM &amp; &lt;100 µM</td>
<td>–</td>
<td>8:176:2:38</td>
<td>0.11</td>
<td>0.69</td>
<td>0.29</td>
<td>0.11 to 0.73</td>
</tr>
<tr>
<td>Single liability</td>
<td>&gt;1 µM &amp; Single liability</td>
<td>–</td>
<td>13:557:50</td>
<td>0.18</td>
<td>0.91</td>
<td>2.28</td>
<td>0.75 to 6.84</td>
</tr>
<tr>
<td></td>
<td>&lt;1 µM &amp; single liability at IC&lt;sub&gt;50&lt;/sub&gt; &lt; 100 µM</td>
<td>–</td>
<td>5:965:46</td>
<td>0.07</td>
<td>0.83</td>
<td>0.39</td>
<td>0.12 to 1.24</td>
</tr>
<tr>
<td>Dual liabilities</td>
<td>&gt;1 µM &amp; dual liabilities at IC&lt;sub&gt;50&lt;/sub&gt; &lt; 100 µM</td>
<td>–</td>
<td>5:065:55</td>
<td>0.07</td>
<td>1</td>
<td>9.3</td>
<td>0.50 to 172.31</td>
</tr>
<tr>
<td></td>
<td>&lt;1 µM &amp; Dual liabilities at IC&lt;sub&gt;50&lt;/sub&gt; &lt; 100 µM</td>
<td>–</td>
<td>2:868:47</td>
<td>0.03</td>
<td>0.85</td>
<td>0.17</td>
<td>0.035 to 0.85</td>
</tr>
<tr>
<td>Triple liabilities</td>
<td>&gt;1 µM &amp; Triple liabilities at IC&lt;sub&gt;50&lt;/sub&gt; &lt; 100 µM</td>
<td>–</td>
<td>11:059:55</td>
<td>0.16</td>
<td>1.00</td>
<td>21.45</td>
<td>1.23 to 372</td>
</tr>
<tr>
<td></td>
<td>&lt;1 µM &amp; Triple liabilities at IC&lt;sub&gt;50&lt;/sub&gt; &lt; 100 µM</td>
<td>–</td>
<td>1:269:53</td>
<td>0.01</td>
<td>0.96</td>
<td>0.38</td>
<td>0.033 to 4.34</td>
</tr>
</tbody>
</table>

The thresholds with high odds ratio for each assay/liability are highlighted.
CI, confidence interval.
instances is >96%. This means compounds falling in these quadrants are most-likely to be most-DILI compounds. A number of most-DILI drugs from the current dataset showed low safety margins (<10) in all 3 mechanistic assays. Examples of these drugs withdrawn from the market for severe DILI are benzamone, benzamorone, troglitazone, etoridone, or drugs that carry black box warnings for hepatotoxicity such as tolcapone, ketoconazole, nefazodone, tipranavir. All of these carry black box warnings for hepatotoxicity such as tolcapone, ketoconazole, nefazodone, tipranavir. All of these drugs have a plasma \( C_{\text{max, total}} > 10 \mu M \) except nefazodone that has a plasma \( C_{\text{max, total}} = 6.8 \mu M \). Although some no-DILI compounds were active in these assays, 19 in the THLE or HepG2 cytotoxicity assays, 8 mitochondrial inhibitors or uncouplers, and 9 BSEP inhibitors all within a potency range between 0.3 and 100 \( \mu M \), their plasma \( C_{\text{max, total}} \) in the majority of cases were <1\( \mu M \) achieving apparently adequate margins of safety generally >100-fold in most cases (Figs. 3B, 3D, and 3F).

Figure 4 shows compounds which are active <100 \( \mu M \) in individual or combinations of these assays along with their plasma \( C_{\text{max, total}} \). The majority of no-DILI compounds, despite having liabilities in one or more assays, have plasma \( C_{\text{max, total}} \) values <1\( \mu M \) whereas a majority of most-DILI compounds with activity have plasma \( C_{\text{max, total}} \) values >1\( \mu M \). Understanding mechanistic liabilities can further segregate no-DILI from most-DILI drugs. In this regard, compounds with triple liabilities and plasma exposure >1\( \mu M \) separated most-DILI-concern compounds from no-DILI-concern compounds with high odds ratio of 21 relative to compounds with single (odds ratio of 2.2) or dual liabilities (odds ratio of 9.3, Table 1). For instance, 11 most-DILI compounds had exposure >1\( \mu M \) and were active in the cytotoxicity, BSEP inhibition, and mitochondrial impairment assays. Only 2 no-DILI compounds (oxybutynin and phenzopyridine) had activities in all 3 assays with corresponding \( C_{\text{max, total}} \) Values <0.2\( \mu M \). The separation for most-DILI-concern from no-DILI-concern compounds with a single liability like cytotoxicity appears to be poor. A large number of most-DILI compounds \( n = 33 \) were inactive in all of these assays up to the maximum testing limit.

Finally, we performed a retrospective analysis of pharmaceuticals outside of the LTKB database that caused transaminase elevations in clinical trials (Phase-I, II, or III) and investigated if this approach would have identified hepatotoxic liabilities based on exposure and mechanistic assessment. Table 2 shows Pfizer discontinued historical compounds (Feng et al., 2009; Shah and Greene, 2014) and more recent compounds from other pharmaceutical companies Takeda’s TAK-875 (Leifke et al., 2012), Lilly’s LY-2409021 (Kelly et al., 2015), Merck’s MK-0893 (Ruddy, 2011), and Addex’s ADX-10059 (Maggos, 2009) that had evidence of liver injury during clinical development. As shown, all compounds except ADX-10059 have total plasma exposure >1\( \mu M \). The majority of these compounds have low safety margins in the mechanistic assays corresponding to what we observed as statistically significant for a subset of compounds from the LTKB (see Table 2). For example, Pfizer’s CP-085958 and TAK-875 had low safety margins for mitochondrial (<1.2-fold) and BSEP inhibition (<1.4-fold) although the IC\(_{50}\) in both assays were >10\( \mu M \). Likewise, Pfizer’s CP-368296 had a low margin for HepG2 cytotoxicity (3-fold) and BSEP inhibition (1.1-fold). CP-724714 had low margin for all 3 assays, cytotoxicity (4.7-fold), mitochondrial inhibition (1.6-fold), and BSEP inhibition (1.8-fold). In addition, LY-2409021 had a safety margin of 0.58-fold in THLE cytotoxicity and 1.9-fold in mitochondrial uncoupling assay. MK-0893 had a low safety margin for mitochondrial inhibition (7.56-fold) besides other cell-based liabilities, mitochondrial uncoupling (21.9-fold), BSEP inhibition (17-fold), cytotoxicity (12-fold). ADX-10059 had multiple liabilities in cytotoxicity (HepG2 IC\(_{50}\) = 27.48\( \mu M \) and BSEP inhibition (IC\(_{50}\) = 25.09\( \mu M \), however its plasma \( C_{\text{max, total}} \) was <0.5\( \mu M \).
resulting in relatively higher margin (~65-fold) compared with other compounds represented here. Our findings demonstrate the opportunity to apply principles learned from analyzing compounds from the LTKB with novel pharmaceuticals.

Finally, it has been shown earlier in the form of “Rule of 2” for hepatotoxicity prediction that total daily dose of 100 mg and above and lipophilicity above 3 are associated with significant risk for DILI using compounds present in LTKB (Chen et al., 2013a). We investigated the rule of 2 in context with mechanistic liabilities (single, dual, or triple) of compounds present in current dataset to examine if their combinations improve sensitivity. As shown in Figure 5, 23 most-DILI compounds and 3 no-DILI compounds from current dataset satisfied the rule of 2 (ClogP > 3 and daily dose ≥100 mg) resulting in sensitivity of 33% and specificity of 95%, better than those shown for single, dual, or triple in vitro assay liabilities in Table 1. However, 21 of 23 of most-DILI concern compounds from the current dataset that satisfied the rule of 2 also had single, dual, or triple assay liabilities. Only 2 of most-DILI compounds were additionally identified by the rule of 2 that were inactive in mechanisms investigated here. Compounds with a higher daily dose combined with high lipophilicity may interact with a number of molecular targets and result in DILI due to multiple versus single mechanism. In fact out of 23 compounds in the current dataset that satisfied the rule of the 2, 9 compounds had triple liabilities and 7 compounds had dual liabilities whereas 5 of them had a single liability.

**DISCUSSION**

A few reports have shown an important association between higher total daily drug dose and increase occurrence of DILI (Chen et al., 2013a; Lammert et al., 2008). In some cases where the mechanism of DILI is strongly related to reactive metabolite formation and covalent binding, the application of normalizing to daily dose appears to be more advantageous for discriminating DILI potential (Usui et al., 2009). For this mechanism, daily dose may better reflect maximum hepatic exposure, and would be a better way to represent the data than systemic exposure according to the authors (Usui et al., 2009). While dose is important, it is meaningless for compounds that are innocuous for causing liver injury and for calculating safety margins for in vitro assays due to variability in drug bioavailability. Hence, we postulated that consideration of systemic exposure would improve DILI prediction in conjunction with mechanistic assays linked to hepatotoxic liabilities.

In theory, the free plasma exposure should correlate with free tissue exposure at least for membrane permeable neutral drugs, and should reasonably correlate with toxicity. However, for the current dataset, a plasma C_{max,total} of >1 μM correlated well with an increased likelihood of DILI, whereas the correlation with plasma C_{max} was poor. This result may partly reflect the influence of active liver transport resulting in free liver concentrations higher than free plasma concentrations, but needs further evaluation. For example, organic anion transporting polypeptide (OATP) family transporters in particular OATP1B1 and OATP1B3 are expressed on the sinusoidal membrane of hepatocytes and facilitate hepatic uptake of certain drugs (Kalikowski and Niemi, 2009). Higher liver exposure would therefore reduce perceived safety margins based on plasma exposure alone. Many statins are known substrates of OATP1B1 and have been reported to have a higher exposure in hepatocytes (Shitara et al., 2013). From the present dataset, most-DILI compounds such as pazopanib and the sulfate metabolite of troglitazone are OATP1B1 substrates; erythromycin and ritonavir are OATP1B3 substrates; bosentan, methothrexate, and rifampin are substrates of both transporters (Shitara et al., 2013). A majority of these drugs have low safety margins in the mechanistic assays investigated here. On the other hand, digoxin and fexofenadine, despite being OATP1B1 and/or OATP1B3 substrates (Shitara et al., 2013), do not cause clinical DILI. The safety margin for cytotoxicity for digoxin using plasma C_{max} was 101-fold, despite a HepG2 cytotoxicity IC_{50} value of 0.35 μM. Fenofoxafedine (plasma C_{max,total} = 0.98 μM) was found to be inactive in all assays evaluated here. Figure 6 shows chemical structures and safety margins of these known OATP substrates from the current dataset.

It may be that for other highly permeable compounds (e.g., lipophilic basic compounds) that undergo more passive liver uptake, consideration of plasma C_{max} may result in a
conservative estimation of the safety margins. Practically, a conservative estimate may be better for drugs intended for treatment of chronic indications or in individuals predisposed to liver injury due to genetic polymorphisms such as the functional expression of liver canalicular BSEP and MRP2 transporters (Meier et al., 2006).

Projections of efficacious human plasma exposure for lead compounds may not always be accurately determined early-on. In such instances, deprioritizing compounds with multiple liabilities may be desirable since they appear to have lower plasma exposures at which liver injury occurs relative to drugs with single or no liabilities (Fig. 4). Advancing drugs with even moderate to weak activities in these assays may increase the odds for DILI if the actual human exposure is higher than estimated or if unanticipated liver uptake or accumulation occurs. For example, the projected efficacious plasma concentration for CP-724,714 at the time of clinical nomination was nearly 10-fold lower than the plasma $C_{\text{max,total}}$ at which hepatic and cholestatic injury occurred in humans. Later studies showed active, as well as passive, uptake of CP-724,714 by human hepatocytes that would further diminish actual margins based on plasma $C_{\text{max,total}}$ (Table 2) (Feng et al., 2009; Woodhead et al., 2014). Although we cannot prove causality with this and other drug candidates based on these known mechanisms of hepatotoxicity, there are numerous drugs like troglitazone that adversely affect mitochondrial function and/or bile salt transport that cause “idosyncratic” DILI (Dawson et al., 2012; Morgan et al., 2010, 2013; Porceddu et al., 2012; Woodhead et al., 2014). Once information on plasma $C_{\text{max,total}}$ of drug becomes available in early human clinical trials, safety margins in these in vitro assays could be used to guide for dose selection in clinical trials.

Only 20% of most-DILI compounds ($n = 14$) in the current dataset have total plasma exposure <1 μM or have high safety margins, including 3 withdrawn drugs (nomifensine, perhexiline, and zimelidine) and 4 drugs with black-box warnings (flu-tamidine, zacitabine, methotrexate, and amiodarone). Eight of these 14 compounds with plasma $C_{\text{max,total}} < 1$ μM were active in BSEP, mitochondrial, or cytotoxicity assays at $IC_{50} < 100$ μM as shown in Figure 4 adding value beyond plasma exposure. The majority of most-DILI drugs presented here including those shown in Table 2 have plasma $C_{\text{max,total}} > 1.1$ μM which further limit our ability to validate the predictivity of current in vitro assays for low plasma $C_{\text{max,total}}$ compounds. However, some of these compounds have other mechanisms causing hepatotoxicity not investigated here. For instance, the triphosphate metabolite of zacitabine impairs mitochondrial DNA polymerase gamma replication at a low multiple of its efficacious concentration in HIV/HCV co-infected patients (Kakuda, 2000). Amphiphilic drugs such as amiodarone ($C_{\text{max,total}} = 0.85$ μM) and perhexiline ($C_{\text{max,total}} = 0.4$ μM) can accumulate further in liver mitochondria or lysosomes due to their basic lipophilic structural properties and interfere with these organelle functions independent of active liver transporter uptake (Fromenty and Pessayre, 1995). Such effects are not typically captured in the mitochondrial assays utilized here due to relatively short drug incubation time of ~15 min. Also, in the present analysis, easily available rat mitochondria are being used instead of human which may not be suited to identify DILI potential of certain drugs such as the nucleoside analogue fialuridine (Lee et al., 2006). Although, the current approach is not all inclusive and may miss narrowly defined mechanisms of hepatotoxicity such as that of fialuridine, it can still be applied to potential pharmaceuticals as shown in this study (Table 2). Another aspect not

FIG. 5. Relationships between total daily dose and lipophilicity for 125 compounds used in this study. The values are shaped by single, dual, or triple liabilities as shown. The vertical dotted line is at 100 mg total daily dose while horizontal line is at ClogP of 3.
addressed in this study is covalent binding of metabolites with hepatic proteins that may produce DILI. Since the liver is the primary site for the bioactivation of drugs, a few drugs like nomifensine and flutamide possess toxicophores and have documented evidence for reactive metabolite formation leading to hepatotoxicity despite submicromolar plasma exposure in humans (Stepan et al., 2011). Since DILI can occur through multiple mechanisms beyond those examined here, lack of activity in these cell-based assays alone, like that observed for CP-456773 and CP-422935, should not be construed as conferring no-DILI risk potential in humans. Even with caveats such as inability to predict immune-mediated DILI and address reactive metabolite-mediated toxicity due to metabolic incompetence of these cell lines, the predictions based on direct cell-based risk factors in assays presented here are compelling.

Beyond plasma exposure and in vitro assays investigated here, improvements in the sensitivity and specificity for DILI prediction can be made by addition of orthogonal mechanisms such as detection of reactive metabolite formation, inhibition of transporters, eg, multidrug resistance-associated protein, understanding impact on nuclear hormone receptors such as Farnesoid X receptor that regulate transporters such as BSEP or using cell lines that selectively express CYP3A4 to understand drug metabolism-mediated cytotoxicity (Greer et al., 2009; Morgan et al., 2013; Rodrigues et al., 2014; Usui et al., 2009). In fact, Thompson et al. (2012) have already shown utility of multifactorial approach for DILI prediction by combining multiple in vitro assays covering some of these mechanisms along with covalent binding to identify idiosyncratic DILI potential of 27 drugs with 100% sensitivity and 95% specificity. Our major
finding here with a larger dataset is that 80% of most-DILI compounds used in this analysis are high exposure compounds and may result in DILI via multiple or independent mechanisms beyond those investigated here. For example, most-DILI compounds such as nefazodone, troglitazone, benzbromarone with triple liabilities in our analysis are also known to form covalent adducts, are positive for cytotoxicity in a THLE cell line selectively expressing CYP3A4, and are inhibitors of multidrug resistance-associated protein 4 (Morgan et al., 2013; Thompson et al., 2012).

Earlier screening assays for hepatotoxicity reported by Pfizer were based on high content screening approaches measuring various cell injury pathways in HepG2 (O’Brien et al., 2006) or human hepatocytes (Xu et al., 2008) with high sensitivity/specificity when measured at 100× a known human plasma C\textsuperscript{max} for a compound. This multi-parameter approach measuring subethical pathways improved sensitivity beyond conventional cytotoxicity assays alone, but are cost prohibitive with large screening needs for early discovery support and medicinal chemistry design requirements based on structure toxicity relationships. The bulk of discovery support in these areas for Pfizer are based on predicting safety characteristics using physicochemical properties and in vitro assays such as the THLE and HepG2 cells (Greene and Song, 2011). We have recognized for some time that accurate prediction of in vivo toxicity, either as a specific end point such as hepatotoxicity or as a general classification system, will require the use of combinations of in vitro assays and molecular properties. Additional cost effective assays of mechanistic relevance (BSEP and mitochondrial inhibition) at this stage of drug discovery are an effective way to build in this intrinsic liability assessment early on using our existing screening platform. Furthermore, the use of simplified assays that highlight intrinsic risks are preferred at this stage since high content screening approaches can miss identification of pathways that are time-dependent, influenced by hermetic responses, and do not necessarily identify a mechanism of toxicity but highlight a cell injury pathway (eg, loss of mitochondrial membrane potential due to sustained generation of reactive oxygen species generated by a compound vs inhibition of mitochondrial activity cause by a compound). High content approaches that are more costly can then be reserved for evaluating new drug candidates that are of sufficient maturity and interest to teams and are being positioned for initial testing in animals when a more accurate estimate of human C\textsuperscript{max} may be applied.

Beyond consideration of systemic exposure and the mechanistic liabilities described herein, it is likely that predictions of DILI potential will be improved with greater emphasis on modeling of broad mechanisms associated with liver injury, modeling of human liver exposure at efficacious doses, and by consideration of total duration of drug exposure. Chronic medications with intrinsic liabilities to affect organelles and transporters, such as mitochondria and BSEP, may eventually trigger liver injury due to prolonged exposure. Although better biomarkers of these effects are needed in preclinical or clinical studies, the DILISym software may better address this issue since it can integrate multiple mechanistic and pharmacokinetic data for parent drug and its metabolites. This in turn has been successfully used to model dose-, time-, and duration of exposure effects on incidence and severity of liver injury due to cumulative risks with various drugs (Howell et al., 2012; Yang et al., 2014).

In summary, this study highlights plasma C\textsuperscript{max,total} as a major driver in separating most-DILI and no-DILI compounds that are withdrawn, discontinued, or currently marketed drugs. Plasma C\textsuperscript{max,total} was found to be more discriminatory than the plasma C\textsuperscript{max,u} for DILI prediction in lieu of having hepatocellular drug concentrations that are believed to be most relevant. This result would benefit from further testing with a validation set of compounds, beyond those found in the LTKB database and recent examples presented in Table 2. Our findings also emphasized lack of experience for low exposure compounds causing severe DILI in the clinic via fundamental mechanisms such as BSEP and mitochondrial impairment (see Figs. 3C and 3E, bottom left quadrant). Future efforts need to be focused on understanding the in vitro assay profile of low exposure compounds (plasma C\textsuperscript{max,total} < 1 mM) causing DILI in the clinic by some of the key mechanisms of hepatotoxicity investigated or highlighted here. In addition, we showed how consideration of plasma C\textsuperscript{max,total} improved the predictivity of mechanistic assays for DILI and provided important margins in these assays that could be effective in separating most-DILI versus no-DILI compounds. We also validated utility of derived margins from the analysis using the LTKB using small set of recent pharmaceuticals with transaminase elevations in the clinic which showed plasma C\textsuperscript{max,total} > 1.1 mM and low safety margins in one or more of mechanistic assays investigated here. A large number of compounds from LTKB were not active in mechanisms investigated here. Considering the low sensitivity (<30% in all instances) based on the safety margin of individual assays, it is desirable to develop a quantitative scoring scheme to predict severe-DILI risk by combining safety margins from multiple mechanistic assays using orthogonal mechanisms including and beyond those investigated here. The prospective use of the described testing paradigm requires robust preclinical projections of plasma or liver exposure in humans. In the absence of such information, DILI risk of potential drug candidates with activity in multiple mechanistic assays should be carefully weighed against criteria such as on-target potency, primary target location (eg, liver vs brain targeting drugs), etc. Drug candidates with multiple in vitro liabilities should be deprioritized, if possible, to decrease the odds for DILI in the clinic. For marketed drugs this study reinforces using lowest possible efficacious dose, especially when multiple risk factors for hepatotoxicity are present.

**SUPPLEMENTARY DATA**

Supplementary data are available online at http://toxsci.oxfordjournals.org.

**ACKNOWLEDGMENTS**

This study was supported by a cross-line partnership through a Drug Safety Research and Development Science & Technology grant and sponsorship from Denise Robinson-Gravatt and Dale Morris. We acknowledge important technical contributions from Rachel Swiss, Paul D. Bonin, Laurie Warren, Scott Allen, and Andrea L. Nilson for their efforts in generating or coordinating in vitro assay data or compiling plasma exposure data. The authors also thank Robert V. Stanton for his constructive comments on the manuscript.

**REFERENCES**


