Graphical Approaches to Evaluate Liver Safety Data in Clinical Trials

Melissa Schultz* and Scott Diegel*

Abstract

Analysis of liver safety data in clinical trials is a key safety challenge for sponsors and regulatory authorities, and for interim reviews by a Data Monitoring Committee (DMC). Standard tables have a limited ability to address this need; the addition of graphical displays makes it easier for the reviewer to quickly absorb meaningful patterns in the data and discern possible safety signals, helping to improve the evaluation of potential liver concerns.

The University of Wisconsin-Madison Statistical Data Analysis Center (SDAC) serves as an independent biostatistics group, reporting analyses of accumulating data from ongoing clinical trials, for review by independent DMCs. Our reports focus on graphical presentations to facilitate comparisons between treatments and over time, and to communicate a large amount of information in a short amount of time.

We present various ways to use graphics to help address clinical liver safety concerns, such as the frequency and maximum levels of elevations of lab tests across treatment groups, possible correlation between changes in different tests, and patient profiles showing time course including lab results, treatment status, and related events.

Key Words: Clinical trial, graphics, liver safety

1. Background

Timely detection and proper assessment of drug-induced liver injury in clinical trials is one of the key safety challenges for both pharmaceutical industry and regulatory authorities. According to a U.S. Food and Drug Administration (FDA) Guidance,[1] drug-induced liver injury (DILI) has been the most frequent single cause of safety-related drug marketing withdrawals for the past 50 years up through the present. Most drugs that cause severe DILI do so infrequently; typical drug development databases with up to a few thousand subjects exposed to a new drug will show no cases. Databases may, however, show signals of a drug's potential for severe DILI.

The term "Hy's Law" is used to describe cases with the following components:[1]

1. The drug causes hepatocellular injury, generally shown by a higher incidence of 3-fold or greater elevations above the upper limit of normal (ULN) of alanine aminotransferase (ALT) or aspartate aminotransferase (AST) than the control drug or placebo.

2. Among subjects with such elevations, often much greater than 3xULN, one or more also show elevation of serum total bilirubin to $>$2xULN, without initial findings of cholestasis.

3. No other reason can be found to explain the combination of increased aminotransferase and bilirubin, such as viral hepatitis, preexisting or acute liver disease, or another drug capable of causing the observed injury.

*Statistical Data Analysis Center, University of Wisconsin–Madison, Department of Biostatistics and Medical Informatics

2. Objectives

A comprehensive safety assessment should make use of all data available: not only standard liver tests such as ALT, AST, and bilirubin, but also patient demographics, medical history, adverse events, and concomitant medications. Time dependence and correlation among liver test results should also be factored in.

Questions that are important to address when analyzing liver safety data include:

- How many subjects exceed certain threshold values across treatment groups?
- How are changes across different liver tests correlated, and how do those correlations differ between treatment groups?
- Are the patterns of change from baseline different between treatment groups?
- Is there evidence of a dose-response relationship?
- What do time profiles of liver test results look like for individual subjects?

In an ongoing clinical trial overseen by a DMC, the committee members have a limited amount of time to review the accumulating trial data and evaluate the safety profile for participating subjects. Analyses in a DMC report need to be presented in a clear and easy-to-understand manner so the data review, meeting discussions, and decision-making can proceed efficiently. Examples of liver data displays that can help address safety considerations in a clinical trial report are presented here.

The primary mode of presentation in SDAC's DMC reports is graphical.[2] The visual presentation allows the reviewer to easily examine the distribution of the data items and characterize the study population at a glance. Treatment comparisons, both at baseline and over time, are easily examined, as are time-related trends in the data.

In a DMC report, the Introduction and figure captions would include information about the data sources and transfer dates, and a brief description of data-handling and analysis conventions used in the report (which is based on interim data with records that may be incomplete or inconsistent).

Data presented in this poster are taken from de-identified analysis datasets based upon the first CDISC Pilot. This was a collaborative pilot project with FDA and industry, undertaken to assess how CDISC-adherent datasets and associated metadata met the needs and expectations of FDA review staff. The datasets and descriptions are available courtesy of CDISC (Clinical Data Interchange Standards Consortium) at www.ctspedia.org.[3]

3. Examples of Summary Displays

Typically in a clinical trial, blood samples are collected for central laboratory assessments at baseline and at specified times during follow-up. Results from any unscheduled or repeated lab tests are also recorded. Upper (ULN) and lower (LLN) limits of normal ranges for each test, in some cases based on sex and/or age of the subject, are generally included in the laboratory data transferred to SDAC.

3.1 Overall Summary

Figure 1 on page 6 displays the percentage of subjects with any post-baseline elevation above the ULN for liver tests (including unscheduled and repeated measurements), with

stacked bars subdivided by the magnitude of the maximum elevation for each subject. The denominators for percentages for each measure indicate the numbers of subjects with any post-baseline test results available for that measure. The lower panel displays an indicator of the number of subjects experiencing particular elevations of clinical concern, at any time during follow-up.

If the trial data show subjects with elevated values, e.g., ALT or AST >3xULN, we typically produce detailed profile plots for these subjects (see Figure 10 on page 14) so the DMC can review their lab values over time and other data more closely.

Included in reports are tables enumerating values shown in most plots (see Figure 2 on page 7). Open Session Reports include displays aggregated across treatment groups, as in Figure 3, for the sponsor or other participants who must remain blinded to all treatment group information.

Figure 4 on page 9 is an example of an eDISH (evaluation of Drug Induced Serious Hepatotoxicity) plot, displaying the maximum post-baseline values of ALT and total bilirubin for each subject (not necessarily concurrent). Axes are scaled as multiples of the ULN for each test. Potential Hy's Law cases can be spotted easily, as points appearing in the upper right quadrant (based on the dashed green reference lines). However, any assessment of the relevance of individual cases is limited; this display does not take into account time sequence of elevations, or other factors.

Modifications have been proposed that use color/size/shading of plot symbols to represent other aspects of a subject's data, such as sequence (which test's maximum value occurred first) and time between the two tests' maximum results.[4] Displays with additional factors incorporated could be more useful in settings where larger numbers of potential cases are expected, e.g., oncology, hepatitis.

3.2 Timing

For each laboratory test in a report, a display like Figure 5 on page 10 summarizes measurements at scheduled visits during a trial. These displays include change from baseline and the percent of subjects with measurements above the ULN (and/or below the LLN, as applicable) at each visit. In the boxplots, the top and bottom edges of the box represent the 25th and 75th percentiles of the data, and the line in the middle of the box indicates the median. The "whiskers" extend to the 5th and 95th percentiles. A plotting symbol marks the mean. These summary statistics for each group, along with minimum and maximum values if desired, are provided in a supporting table, as in Figure 2.

Unplanned laboratory measurements can arise during a clinical trial for a variety of reasons. The clinical investigator sometimes orders a repeat of a laboratory test, especially if the scheduled test gave an abnormal or unexpected value. The investigator may also request the subject return for a follow-up visit due to clinical concerns. If retests are conducted until the desired results have been reached, analyses from baseline to last observation would be biased toward normality. Thus we generally include only measurements from planned visits when creating displays by visit over time. However, we include all planned and unplanned measurements for analyses that focus on outliers or shifts across the entire study period, such as Figure 1, which uses the most extreme values.

The timing of liver test elevations may also be visualized with a Kaplan-Meier plot of time to first elevation, for a single test or, as shown in this example (Figure 6 on page 11), to the first occurrence of either ALT or AST exceeding 3xULN. In this analysis, follow-up time for subjects with no elevation is censored at the date of their last available test result.

3.3 Shift from Baseline

Figure 7 on page 12 displays a scatter plot of each subject's baseline value and maximum post-baseline value, for ALT and AST in adjacent panels. Treatment groups are identified by different colors and symbols (with the two doses of the active treatment combined in this display). For each test the ULN is not a single number, but takes on different values according to subject characteristics. The grey bars in each plot represent the range of ULN values (for ALT, 32-43 U/L; for AST, 34-36 U/L). The diagonal line indicates equality; points above this line represent a follow-up value greater than the subjects baseline result. Figure 8 on page 13 presents a standard shift table for categorized change from baseline in ALT. Are high follow-up values likely to be associated with high baseline values? Is there a similar distribution among treatment groups? Among different liver tests?

Figure 9 on page 14 presents boxplots for the change from baseline to the maximum post-baseline value, with multiple tests displayed side-by-side with a common y-axis. Do ALT and/or AST demonstrate more pronounced elevations from baseline with active treatment than placebo? Is there a consistent trend evident among the tests? Is there evidence of a dose response?

3.4 Patient Profile Plots

The patient profile plots in Figure 10 on page 14 provide sample displays of longitudinal information for subjects with elevated ALT or AST values. All recorded liver test values are provided as well as study status, dosing information and adverse events (AEs), in addition to pertinent demographic and baseline data. These plot types are easily customizable to display a wide variety of information presented over time, and its relation to dosing and clinical events.

4. Additional Information on SDAC Reports

- Data processed with SAS, graphics created using R
- Focus on graphical presentation facilitates review and interpretation of data
- Integration of data from different sources (lab, CRF, IVRS, SAE) provides a valuable overview of potential toxicity / safety issues
- Flexibility in reporting system allows for ongoing modifications and additions if needed over the course of a trial, in response to evolving trends, questions from DMC members, or external concerns
- Balance between aggregate summary graphics and detailed displays of individual subject data allows for both general treatment group comparisons and close review of specific cases
- Annotations on each graphical page include N's, p-values, identification of data source(s) used, and a cross-reference link to a supporting back-up table (clickable hyperlink in the PDF file)
- We produce in parallel a Closed Session Report (by treatment group) for DMC review only and an Open Session Report (aggregate data only) for the sponsor or other participants in an Open Session of a DMC meeting

- In a Closed Session Report, treatment groups are indicated by letter codes (“A”, “B”, “C”). The assignment of codes to study drug arms is generally not identified in the report, but will be provided to DMC members upon request (either verbally or in writing, as stipulated in a DMC Charter document or otherwise agreed to prior to data review).

- With >2 treatment groups, a choice of contrasts is available for analysis and display (all pairwise comparisons; each active dose group individually vs. control; active dose groups combined vs. control) depending on specifics in the study protocol and analysis plan.

REFERENCES

1. US Department of Health and Human Services, FDA, CDER, CBER (July 2009). Guidance for Industry. Drug-induced Liver Injury: Premarketing Clinical Evaluation. http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM174090
2. SDAC Sample DMC Report. https://www.biostat.wisc.edu/content/sdac-sample-dmc-reports
3. CDISC - FDA Pilot Data Sets. https://www.ctspedia.org/do/view/CTSpedia/DataSetPageCDISC
4. Merz, M., et al. (2014), “Methodology to Assess Clinical Liver Safety Data,” *Drug Safety*, 37 (Suppl 1), S33-S45.

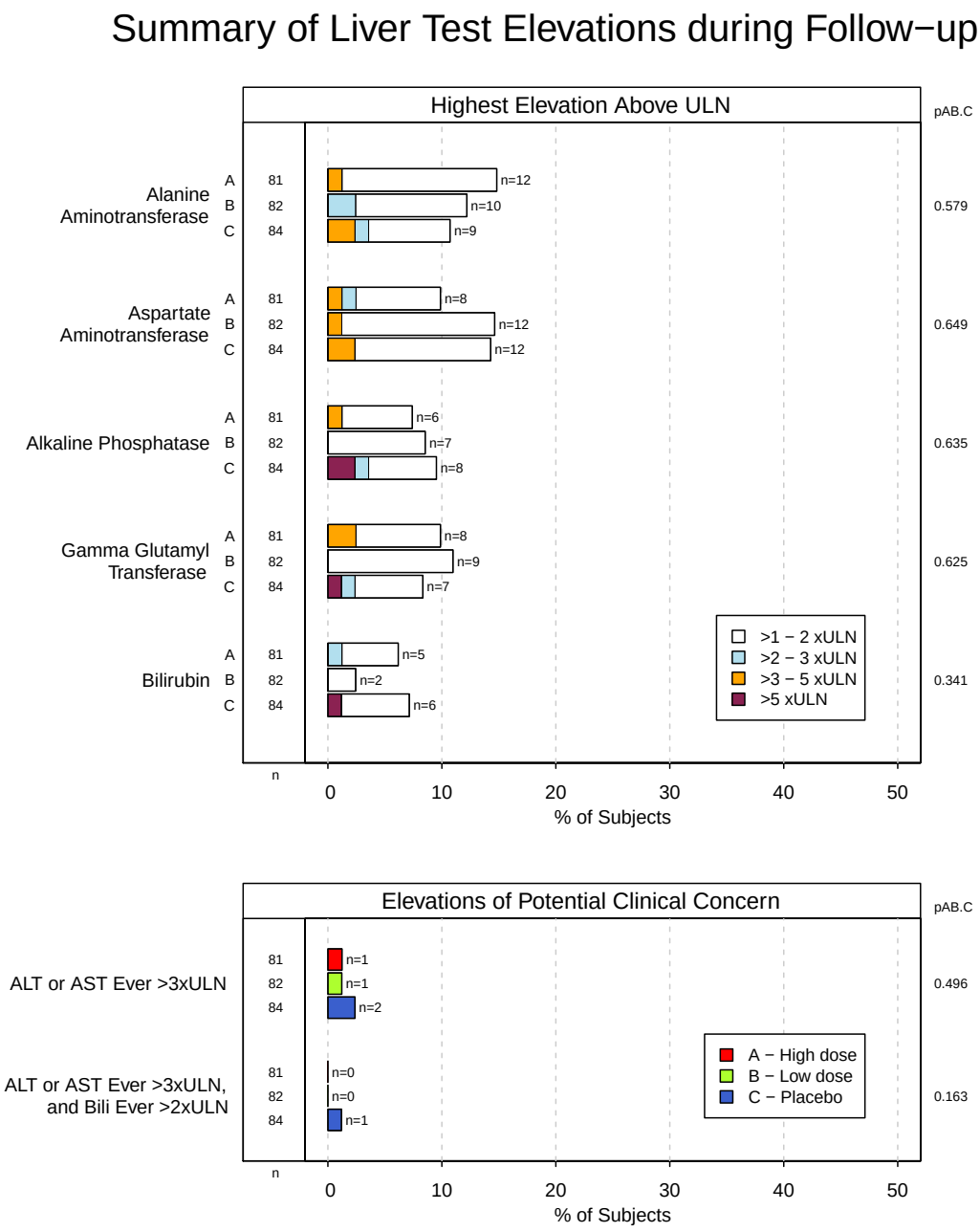

Figure 1: Graphic summary of liver test elevations during follow-up.

Summary of Liver Test Elevations during Follow-up: Highest Elevation Above ULN *(Cumulative Across Categories)*

		Treatment Group									
		A		B		C		TOTAL			P-
	Cumulative N / %	N	%	N	%	N	%	N	%	Contrast	Value
Alanine Aminotransferase	Total Subjs	81		82		84		247		AB.C	0.579
	>5 xULN	0	0.00	0	0.00	0	0.00	0	0.00		
	>3 xULN	1	1.23	0	0.00	2	2.38	3	1.21		
	>2 xULN	1	1.23	2	2.44	3	3.57	6	2.43		
	>1 xULN	12	14.81	10	12.20	9	10.71	31	12.55		
Aspartate Aminotransferase	Total Subjs	81		82		84		247		AB.C	0.649
	>5 xULN	0	0.00	0	0.00	0	0.00	0	0.00		
	>3 xULN	1	1.23	1	1.22	2	2.38	4	1.62		
	>2 xULN	2	2.47	1	1.22	2	2.38	5	2.02		
	>1 xULN	8	9.88	12	14.63	12	14.29	32	12.96		
Alkaline Phosphatase	Total Subjs	81		82		84		247		AB.C	0.635
	>5 xULN	0	0.00	0	0.00	2	2.38	2	0.81		
	>3 xULN	1	1.23	0	0.00	2	2.38	3	1.21		
	>2 xULN	1	1.23	0	0.00	3	3.57	4	1.62		
	>1 xULN	6	7.41	7	8.54	8	9.52	21	8.50		

Note: N and % for each category are cumulative, representing all subjects with that level or higher

Figure 2: Supporting table for summary of liver test elevations during follow-up.

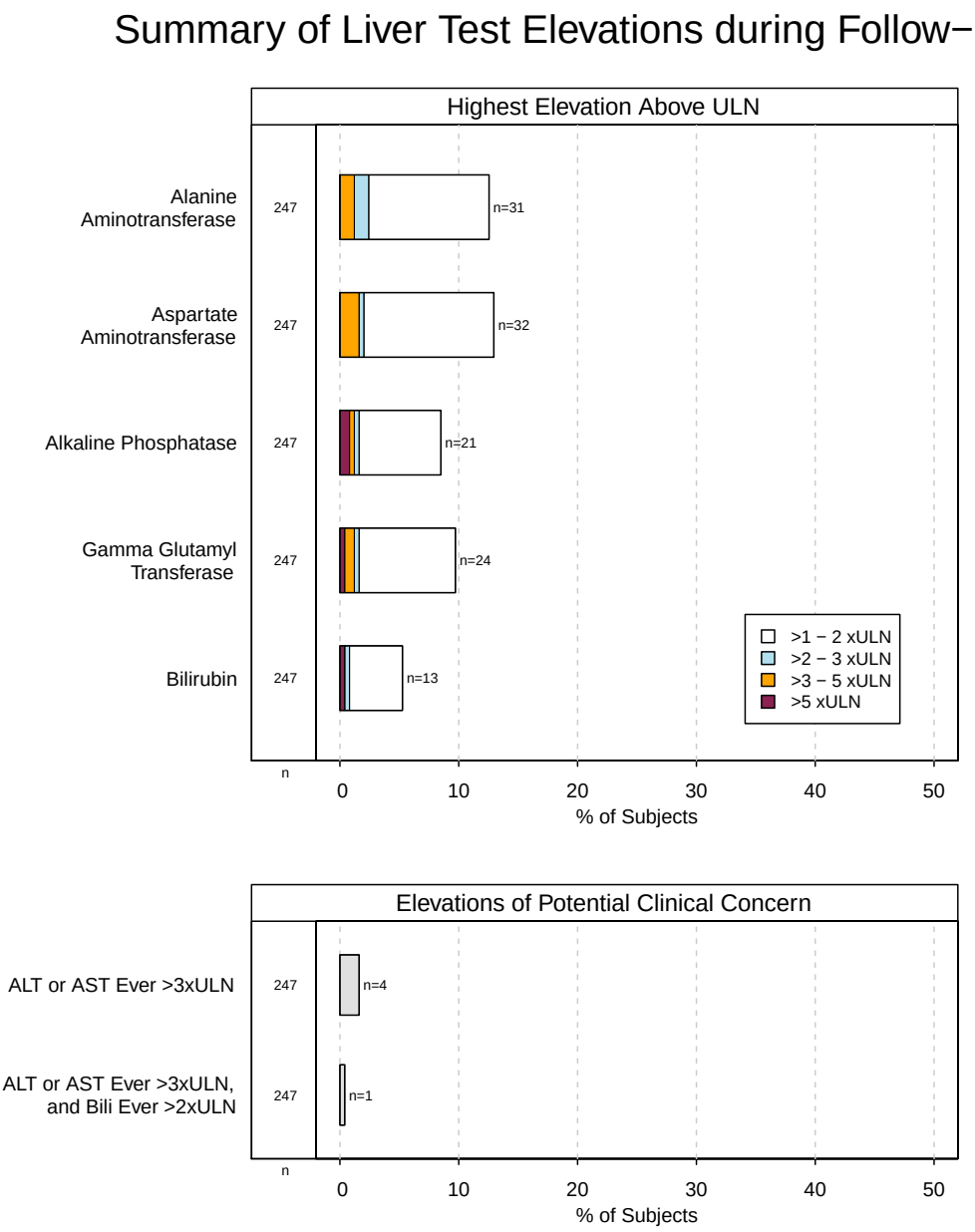

Figure 3: Open Session Report blinded display with aggregate data.

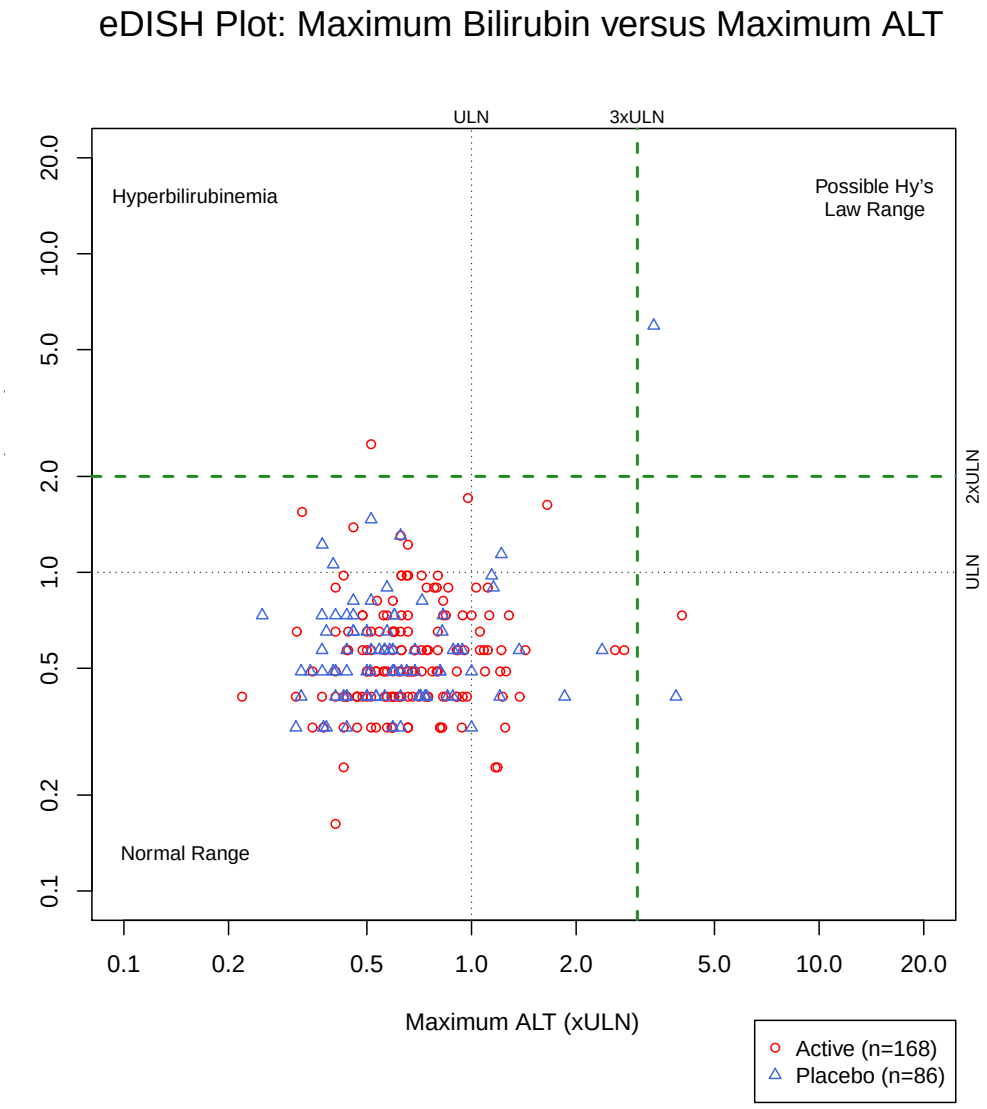

Figure 4: eDISH plot showing the maximum post-baseline values of ALT and bilirubin for each subject.

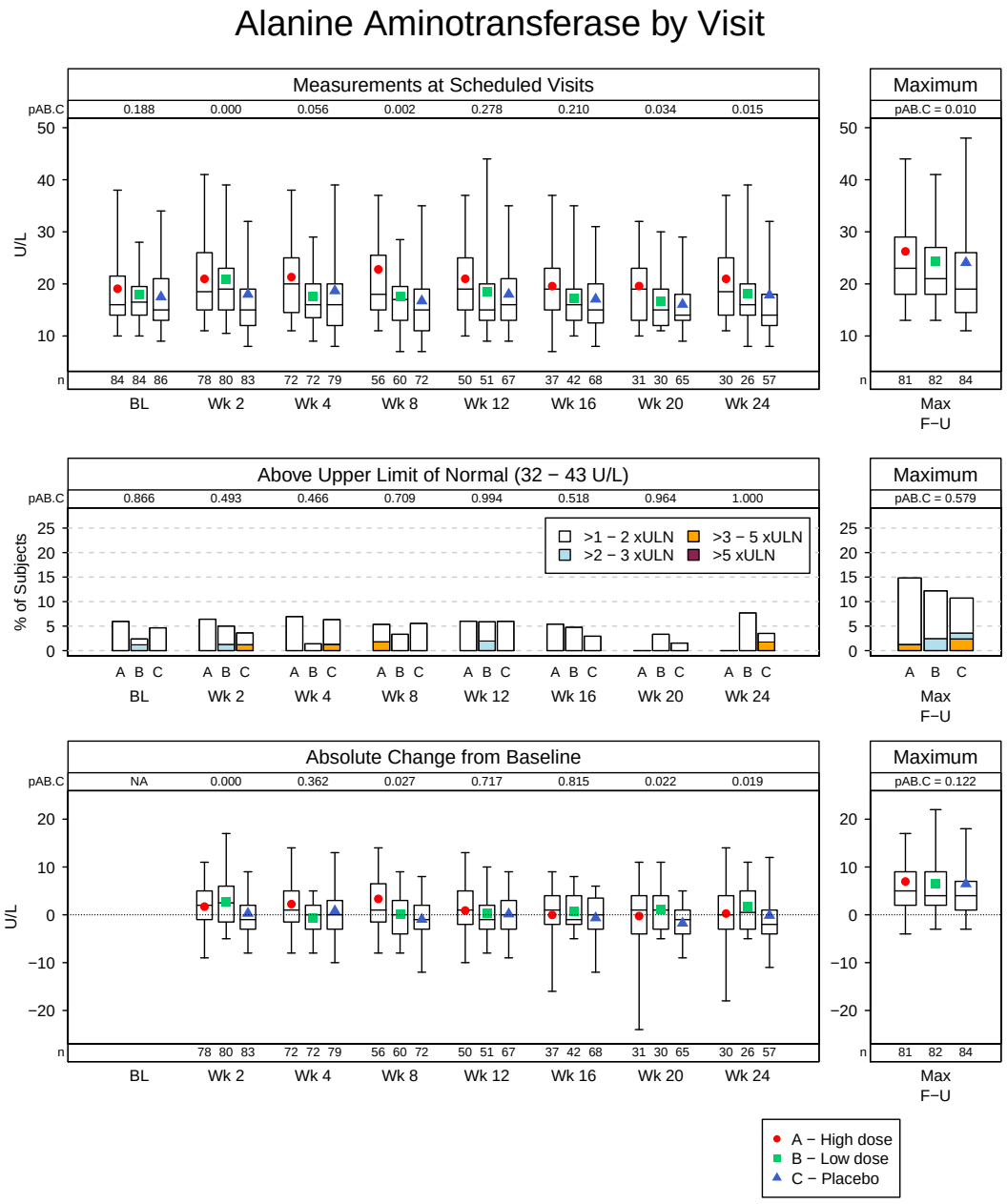

Figure 5: Laboratory measurements at scheduled visits.

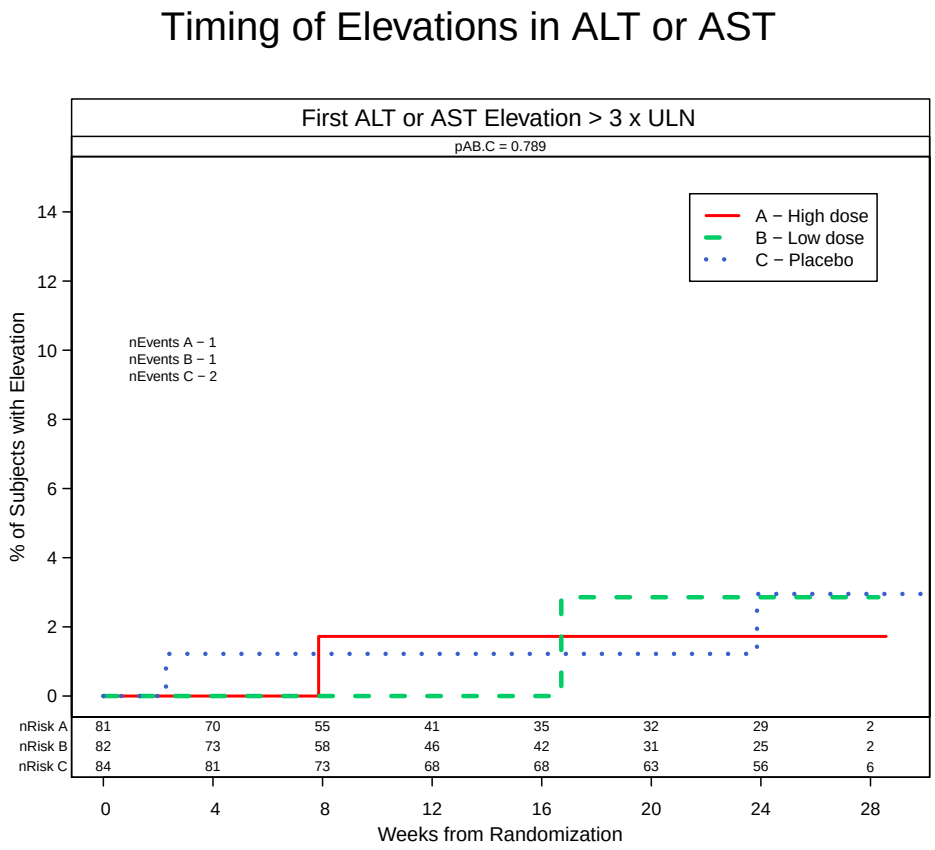

Figure 6: Kaplan-Meier plot of time to first elevation of ALT or AST >3xULN.

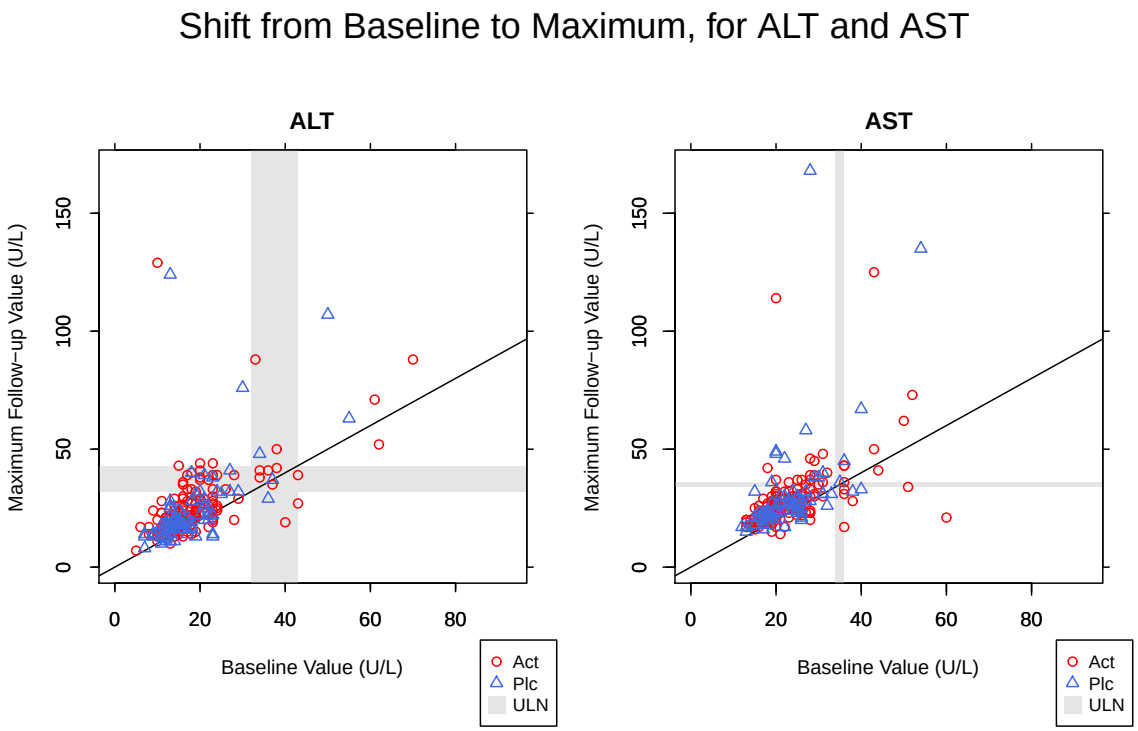

Figure 7: Scatter plots of shift from baseline to maximum value for ALT and AST.

Shift Summary Tables to Assess High Values: Alanine Aminotransferase

		Maximum Post-Baseline Result		
Treatment	**Baseline Result**	Low n (%)	Normal n (%)	High n (%)
High dose (N = 81)	Low	0 (0%)	0 (0%)	0 (0%)
	Normal	0 (0%)	69 (85.2%)	7 (8.6%)
	High	0 (0%)	0 (0%)	5 (6.2%)
Low dose (N = 82)	Low	0 (0%)	1 (1.2%)	0 (0%)
	Normal	0 (0%)	71 (86.6%)	8 (9.8%)
	High	0 (0%)	0 (0%)	2 (2.4%)
Placebo (N = 84)	Low	0 (0%)	0 (0%)	0 (0%)
	Normal	0 (0%)	74 (88.1%)	6 (7.1%)
	High	0 (0%)	1 (1.2%)	3 (3.6%)

Treatment	Shift from Normal/Low to High			
	N	n	%	P-value*
High dose	76	7	9.2	0.70
Low dose	80	8	10.0	0.58
Placebo	80	6	7.5	

N = number of subjects with baseline result “Normal” or “Low” and at least one post-baseline value
n = among the “N”, number of subjects with maximum post-baseline value “High”
* P-values from Chi-square test, each active treatment arm compared with placebo

Figure 8: Shift table for categorized change from baseline in ALT.

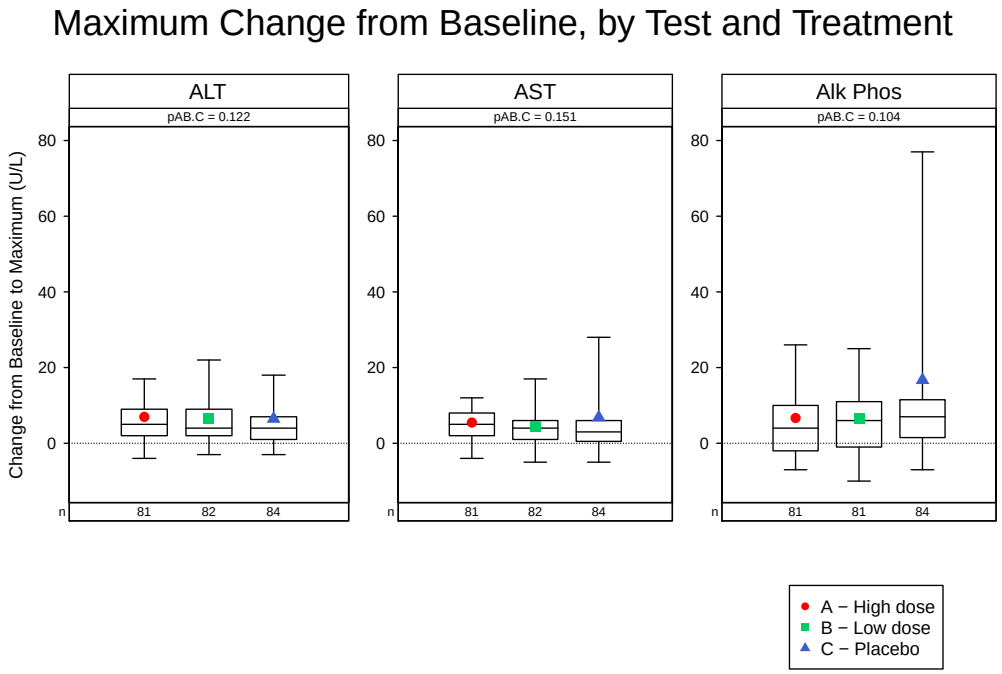

Figure 9: Maximum change from baseline for selected liver safety tests.

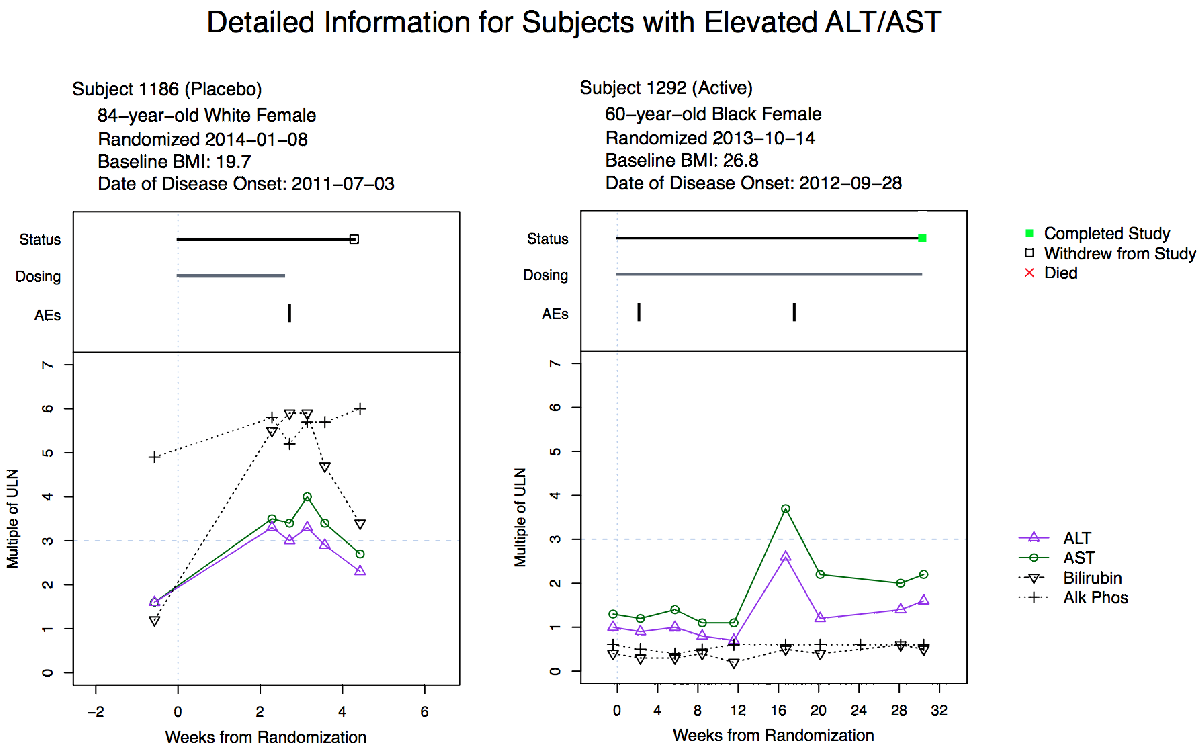

Figure 10: Detailed patient profiles for subjects with elevated ALT or AST.