Chapter 17
Graphics for Safety Analysis

Peter W. Lane and Ohad Amit

Abstract We present a range of graphics designed for reporting the analysis of safety data. For adverse events (AEs), we show a comparative dot-and-interval plot of all the main AEs in a trial and also comparative plots of cumulative incidence and hazard rate for individual AEs. For laboratory data, we show a scatterplot designed to help identify potential liver toxicity and 2 trellis plots that can show several laboratory measurements in a single graph, showing changes from baseline or the relationship between the measurements. We also give an example of a profile plot for individual patients. Finally, for ECG data we show a comparative cumulative distribution plot and a comparative boxplot profile showing how distributions change over time. We also show a simple comparative profile plot of means of an ECG measurement over time. We produced each graph using one of GenStat™, SAS™ and S-PLUS™ (using code very similar to R), as indicated in the text, and the programs and data are available from the Web site associated with the book.

17.1 Introduction

The analysis of safety data mostly takes the form of simple descriptive statistics, displayed in a tabular or graphical form. For example, the number and percentage of patients experiencing adverse events may be presented, or the means or medians of

P.W. Lane (✉)
Quantitative Sciences, GlaxoSmithKine, Stevenage, UK
e-mail: peter.w.lane@gsk.com; peterwlane@gmail.com

O. Amit
Quantitative Sciences, GlaxoSmithKine, 1250 S. Collegeville Road,
Collegeville, PA 19426-0989, USA
e-mail: ohad.amit@gsk.com

clinical laboratory measurements. Graphs are ideal for communicating this type of information concisely. A particular advantage over tabulation is that the descriptive statistics can often be presented in conjunction with the patient data that have been summarized, to put the statistics in context. Another advantage is that the human eye is able to detect anomalies and patterns in pictures better than in tables of numbers, and a graphical display allows more effective communication.

We consider here 3 types of safety information: adverse events, liver toxicity and cardiac safety. We illustrate graphical methods of displaying this information, based on work by a GlaxoSmithKline team of which we were members, and which was reported in a paper by Amit et al. (2008). We have updated these in the light of recent developments and added some new examples. There are clearly many other types of safety information, but we suggest that many may be displayed graphically using the same approach as we use here for the types on which we concentrate.

For each graph in this chapter we present a version drawn by one package (named in the text). The data and code can be found on the website associated with this book. Most of the data are from a single anonymized clinical trial, which we will refer to as the Safety trial.

17.2 Adverse Events

There are numerous adverse events (AEs) reported in each clinical trial, so displays need to be tailored to highlight important information, such as the most common events and events of special interest. The SPERT (Safety Planning, Evaluation and Reporting Team) have recommended a three-tier approach for signal-detection and analysis of AEs (Crowe et al. 2009). The first tier is made up of AEs of special interest, identified in advance of running trials, and the second and third tiers of other AEs that are considered common and uncommon, respectively: "common" is suggested by Crowe et al. to be more than about 1% incidence in any treatment arm, though this will depend on the size of the trial. The methods in this section are suitable for AEs in Tiers 1 and 2; AEs in Tier 3 are best reported with simple summary statistics.

17.2.1 Dot-and-Interval Plot of AE Incidence

A dotplot is an ideal display to show and compare AE incidence in a randomized clinical trial. This type of display was introduced by Cleveland in the context of showing counts and proportions, and is generally considered superior to barcharts and piecharts (Cleveland 1993). Figure 17.1 shows a two-panel display which enhances the simpler dotplot by adding statistical information comparing the incidence rates of AEs in the Safety trial. In this example, we display all AEs that
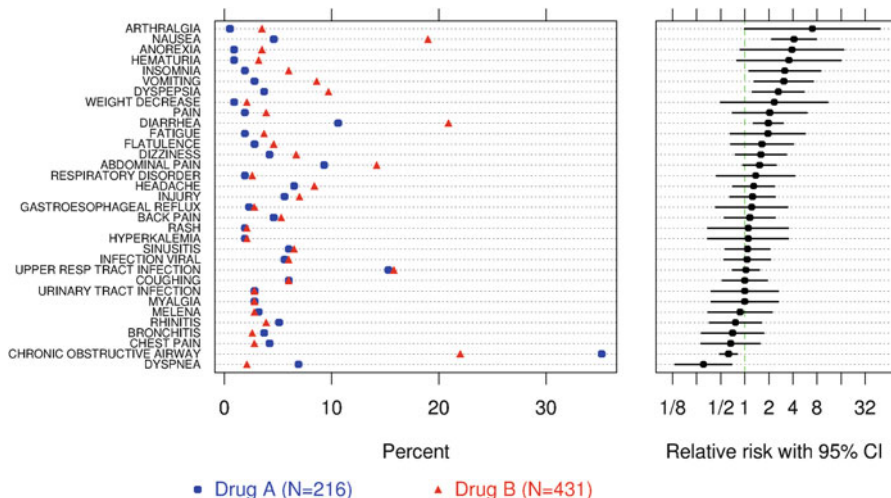
Fig. 17.1 Dot-and-interval plot of AE incidence

had overall incidence greater than 2%, along with relative risks and the asymptotic confidence intervals.

It is clinically valuable to see the actual risk differences as well as relative risks in a single snapshot, to put the statistical ratios into context. Note that the adverse events are ordered by relative risk. Other statistics can also be considered for the right-hand panel, such as risk differences, odds ratios or hazard ratios, depending on the objectives of the display for its audience. It is useful to give further context by adding information about the number of patients in the safety population of the trial: here, we have added that to the key. Colour is used in a modest way to help distinguish the 2 treatments, but note that different symbols are used as well in case the graph is viewed in black and white.

This graph was drawn using S-PLUS™.

17.2.2   Cumulative Incidence of an AE

Cumulative incidence over time is often of interest with AEs, as the time at which such events manifest themselves can be critical in guiding regulators and prescribers regarding monitoring and clinical use of a drug. Figure 17.2 shows a cumulative incidence plot of the gastrointestinal AEs from the same trial as above. It is constructed in much the same way as a Kaplan–Meier (KM) plot, taking account of censored information because many patients withdrew from this trial. It is better to display the information as (1–"survival") against time here, rather than as survival
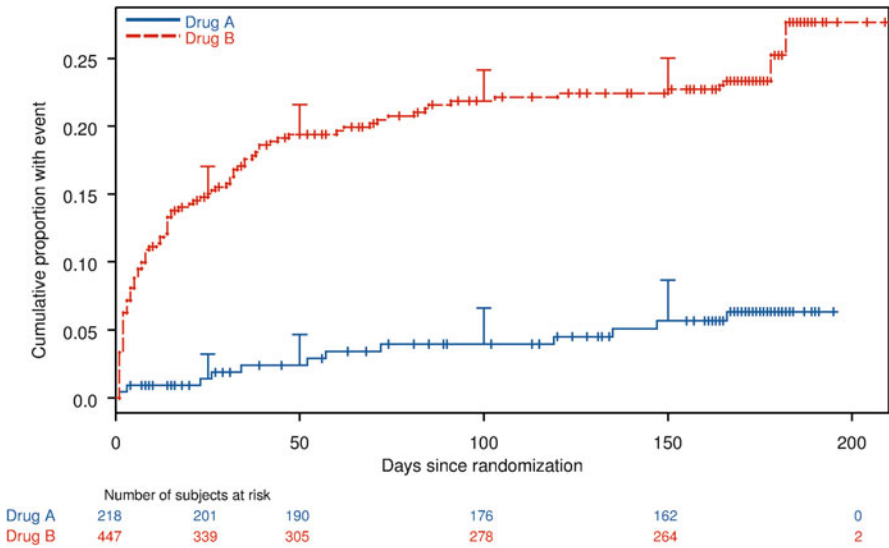
Fig. 17.2 Cumulative incidence of an adverse event, with SEs at selected time-points

against time as in the KM plot because incidence is the focus for AEs. There is also the point that most people are used to seeing survival plots with the *y*-axis ranging fully from 0 to 1, which would cramp the information at the top of the frame (Pocock et al. 2002).

The plot has several enhancements compared to a simple KM plot. First, the numbers of subjects at risk are displayed as strategic points along the *x*-axis in a lower margin, to quantify the steadily decreasing population as subjects withdraw over time. Second, the actual censoring times of subjects on each arm are marked as a "rug-plot" on top of each step function representing the cumulative proportion. Third, the SEs of the estimated proportions are indicated at the same strategic points to show how much precision has been achieved. In this case, we have shown these as positive error bars only, but they could alternatively be negative bars, the usual two-sided bars, or indeed show 95% confidence intervals instead. Note that colour is used as in Fig. 17.1, and that different line styles are used in case the graph is viewed in black and white. In some settings, such as clinical trials without a fixed follow-up, use of competing risks methodology (Pintilie 2007) should be considered in order to estimate the cumulative incidence curves. This methodology would be particularly useful in trials where subjects are treated until the occurrence of a specific event (e.g., disease progression), separate from the safety event of interest.

This graph was drawn using S-PLUS™, which provides an option to add the tricky part, i.e., the rug-plots.
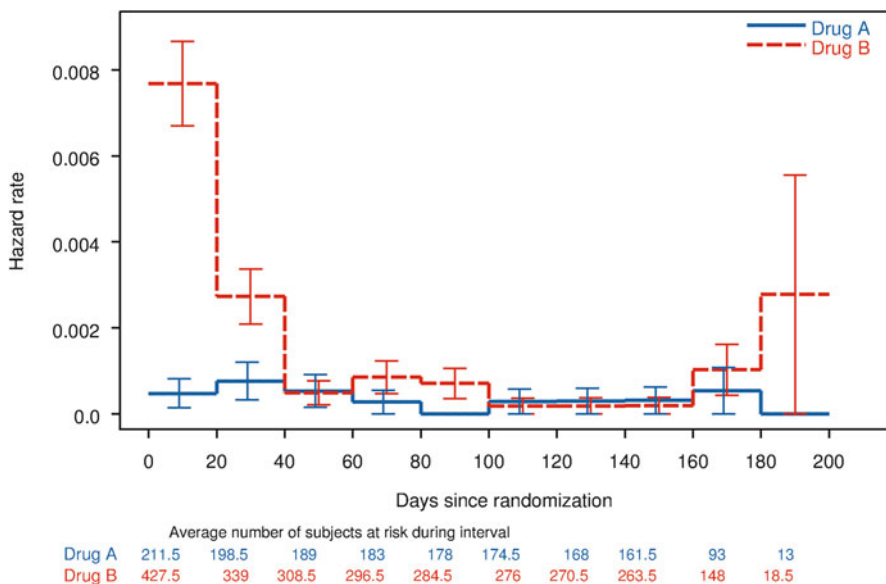

Average number of subjects at risk during interval

Drug A	211.5	198.5	189	183	178	174.5	168	161.5	93	13
Drug B	427.5	339	308.5	296.5	284.5	276	270.5	263.5	148	18.5

Fig. 17.3 Comparative hazard function for gastrointestinal AEs of concern: nausea, abdominal pain, diarrhoea and vomiting

17.2.3   Hazard Rate for an AE

The information about incidence of an AE can also be displayed as a hazard rate function. Figure 17.3 shows this for the same data, with the hazard rates estimated in successive 20-day intervals (again taking account of censoring) and drawn as a pair of step functions.

As for Fig. 17.2, this has been enhanced with a lower margin giving the average number of subjects at risk during each time period, and SE bars for each hazard estimate. The choice of time periods can be important to illustrate the differences between the two drugs effectively. Note that there is no SE for periods where the hazard was 0, as the estimate of SE is formally 0 for such periods. The lines are again differentiated both by colour and line-style.

This graph was drawn with S-PLUS™.

17.3   Liver Toxicity

Drug-Induced Liver Injury (DILI) is "the single most common adverse effect that can result in failure to obtain regulatory approval to market a new drug, and post-marketing regulatory actions include labelling restrictions and withdrawal from the

marketplace" (Watkins 2005). In general, hepatic safety is the second most common reason for termination due to safety during drug development. DILI is the most frequent cause of acute liver failure in patients evaluated for liver transplantation. There are 3 main types of liver toxicity that may be observed: directly destructive, indirect (or metabolic) and cholestatic.

Intrinsic or direct liver injury (e.g., that seen with acetaminophen) is:

- Predictable
- Dose-related
- Similar in animals
- Relatively common
- Observed after a short interval

    On the other hand, idiosyncratic liver injury (e.g., that seen with Troglitazone) is:

- Unpredictable
- Often dose-independent (Lammert et al. 2008)
- Not seen in animals
- Relatively rare: 1 in 10,000 to 1 in 100,000
- Usually observed after a longer interval

    There are 2 main types: hypersensitivity and metabolic.

    Typically in clinical trials, 4 cardinal variables are monitored for liver toxicity using what are described as liver function tests (LFT):

- ALT: alanine aminotransferase
- TBL: total bilirubin
- AST: aspartate aminotransferase
- ALKP: alkaline phosphatase

ALT, AST and TBL are of particular interest because of a criterion that is generally accepted as a surrogate for potential DILI, known as Hy's Law. While there are several clinical aspects to the determination of a Hy's Law case, the laboratory criteria are defined as an elevation of ALT or AST together with simultaneous or subsequent elevation of bilirubin. An occurrence of such a simultaneous elevation indicates the potential for severe liver injury and which in turn could predict for acute liver failure. Andrade et al (2005) have reported a 10% fatality from drug-induced liver injury with jaundice.

17.3.1   Scatterplot to Assess Drug-Induced Liver Injury

The FDA has adopted a criterion for Hy's Law (Wilke et al. 2007), generating a signal when the following conditions are all met:

- ALT or AST≥3xULN (upper limit of normal measurements)
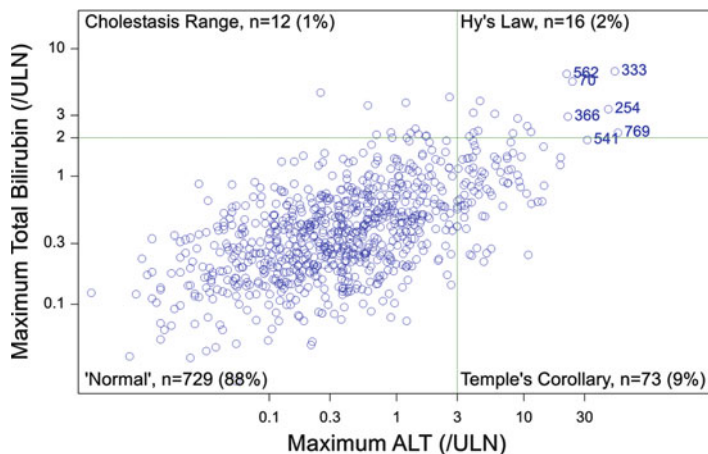- TBL≥2xULN
- ALKP≤2xULN

Fig. 17.4 Scatterplot of Total Bilirubin versus ALT used as a signal for DILI

A graphical approach suggested as part of the FDA DILI guidance to evaluate potential Hy's law cases is shown in Fig. 17.4 (using simulated data). This is a simple scatterplot of maximum TBL for subjects during the course of a trial against maximum ALT, with reference lines and annotation associated with the Hy's Law criterion. This concentrates just on ALT and TBL, and a similar graph can be drawn for TBL versus AST.

The figure is split into 4 quadrants with the upper right quadrant indicating the potential for a Hy's Law case. The bottom right quadrant showing subjects with elevated ALT but without elevated TBL is associated with another conjecture called Temple's Corollary. It has been hypothesized that a significant number of patients within this quadrant will predict for the presence of a Hy's Law case at some point in time. The top left quadrant is referred to as the Cholestasis Range, associated with Gilbert's Syndrome (high TBL but normal ALT).

The aspect ratio of this graph is worth noting: it emphasizes the ALT measurements by having a longer axis, and this corresponds to the fact that one can observe far more extreme ALT measurements as multiples of ULN than with TBL. In addition, ALT is typically more predictive of clinical harm in DILI because TBL outliers may be due to Gilbert's Syndrome and therefore not so important.

This graph was drawn using GenStat™.

17.3.2   Scatterplot Trellis of Shifts from Baseline Measurements

A standard tabular summary of LFTs that evaluates shifts in individual LFT measurements is shown in Table 17.1. The number of subjects who "shifted" to a higher LFT value relative to their baseline value is shown in the table.

Table 17.1 A standard tabular summary of LFTs, evaluating shifts in individual LFT measurements

Test	Time	n	Treatment A			n	Treatment B		
			Any increase	Increase >3xULN	Increase >5xULN		Any increase	Increase >3xULN	Increase >5xULN
ALT	Week 4	18	2 (11%)	1 (6%)	0	18	2 (11%)	1 (6%)	0
	Week 6	18	2 (11%)	1 (6%)	0	18	2 (11%)	1 (6%)	0
	Week 8	18	2 (11%)	1 (6%)	0	18	2 (11%)	1 (6%)	0
	Post Rx	18	3 (17%)	1 (6%)	0	18	3 (17%)	1 (6%)	0
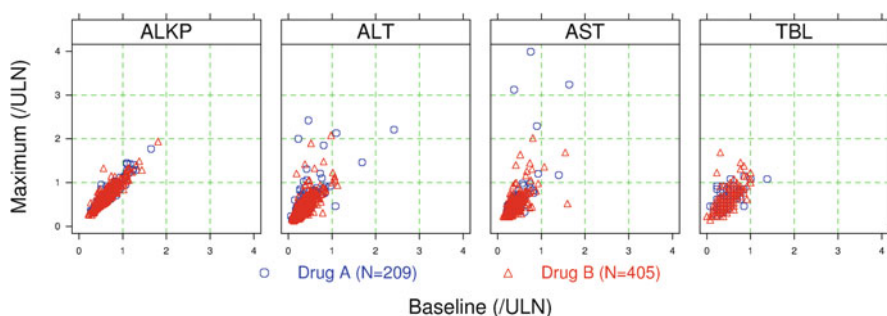
Fig. 17.5 Trellis of scatterplots of maximum LFT measurements versus baseline

A concise graphical summary of the same sort information from the Safety trial is provided in Fig. 17.5.

There are several important elements to this graph. First, when comparing an active drug against a control, as here, it can be more informative to arrange that the control points (blue circles) are drawn last: the distribution of extreme points from the active drug then appears as a "frill" around the central mass of blue points, allowing quick visual appreciation of the potential effect of the drug. However, some software makes this difficult, and we could not find a way to achieve it in Fig. 17.5 using S-PLUS™.

Interpretation of the absolute distribution of the observations for each treatment can be misleading, as the points tend to lie above and to the left of the centre diagonal of each graph. This is an inevitable consequence of using a maximum of several values on the *y*-axis: in this case there were 8 visits during the trial. Because of natural variation, the distribution of a maximum of 8 observations is inevitably shifted upwards compared to the distribution of a single (baseline) measurement, regardless of any effect of the drugs. Note that the reference lines have been updated since publication of this figure in Amit et al. (2008) to take account of the criteria described in the FDA DILI guidance.

Two other features of this graph can be of importance for interpretation. The distribution of LFT measurements is usually very skewed, particularly when the
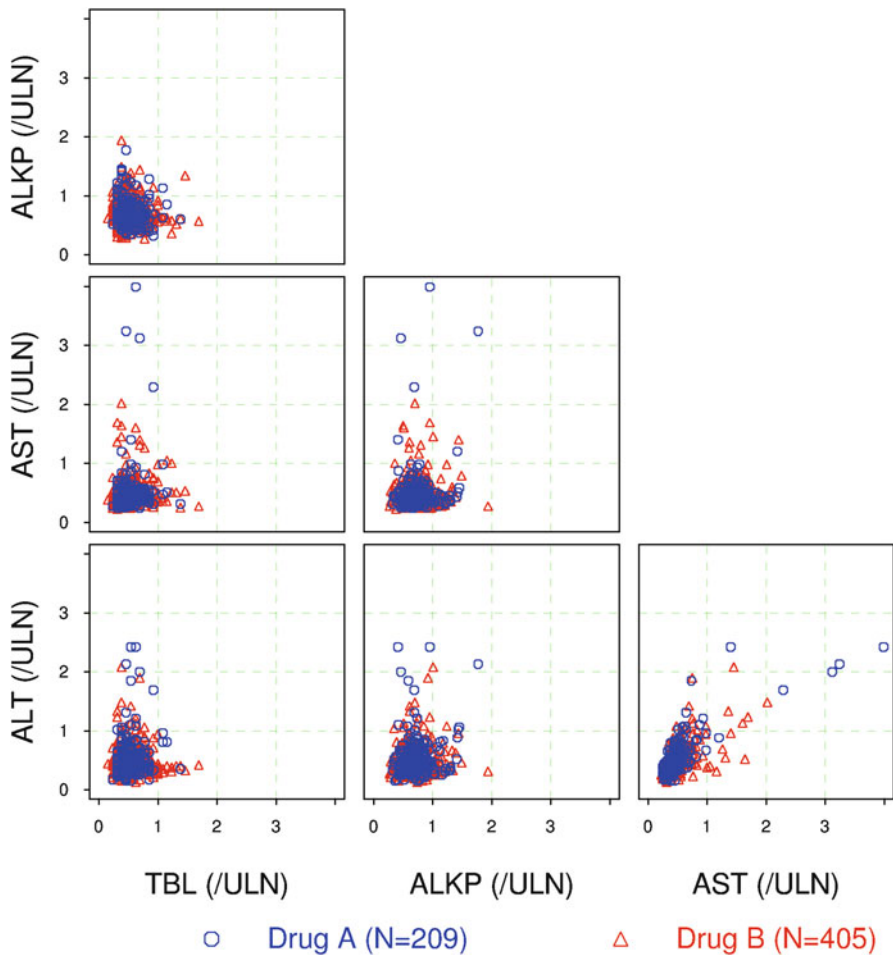
Fig. 17.6 Triangular scatterplot matrix of maximum LFT measurements

patient population has significant elevations, as seen in Fig. 17.4. The graphical display of the relationships can then often be improved by using a log scale as in that figure, which allows display of all the extreme values without overemphasizing those values within the figure.

17.3.3   Scatterplot Matrix of Maximum LFT Measurements

The association between the various LFT measurements can be displayed in the matrix plot as shown in Fig. 17.6. This shows a triangular array of each of 4 LFTs against each other, allowing quick visual assessment of the interrelated information
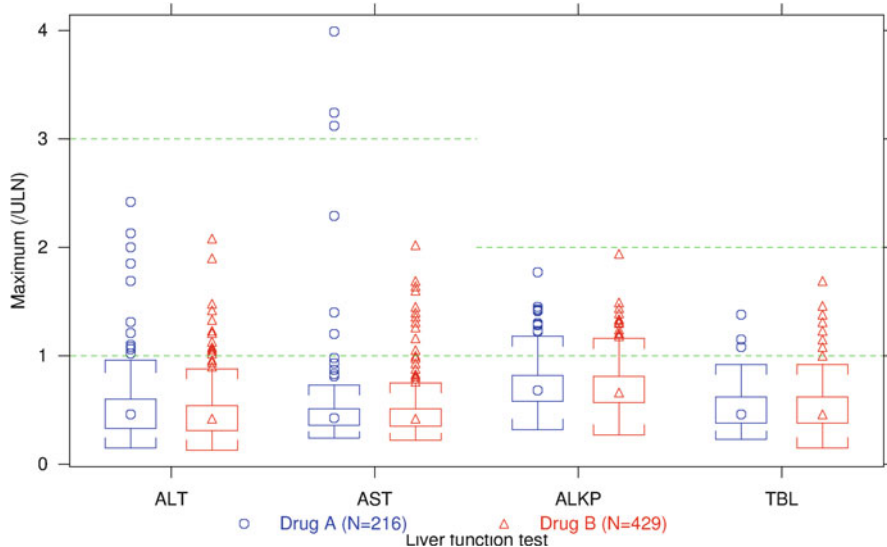
Fig. 17.7 Parallel boxplots of LFT measurements

associated with signals such as Hy's Law. Like Fig. 17.4, Fig. 17.6 can also provide a quick visual assessment of potential Hy's law cases and can also show another important relationship between ALT and AST. The latter can help in further defining the nature of the liver signal of a particular compound.

The individual scatterplots are designed much the same as those in Fig. 17.5. Here, however, we succeeded in arranging for S-PLUS™ to plot the control treatment, Drug A, on top, so that any differences in distribution for Drug B appears as a fringe around that for Drug A.

17.3.4 Parallel Boxplot of LFT Measurements

If the shift information is not of particular interest, and the association between different measures is not to be concentrated on, a simpler graph can give a visual report of the distributions. Figure 17.7 uses boxplots to display the distributions, with an emphasis on the outlying points—a key feature of boxplots.

These boxplots are those defined as "schematic diagrams" by Tukey (1977), with the whiskers extending outside the box no further than 1.5 times the box width. The extreme points are all individually marked, which is ideal for safety measurements of this kind where the interest focuses on them.

This graph was drawn using S-PLUS™, requiring the definition of a transposition function in order to be able to orient the boxes vertically.
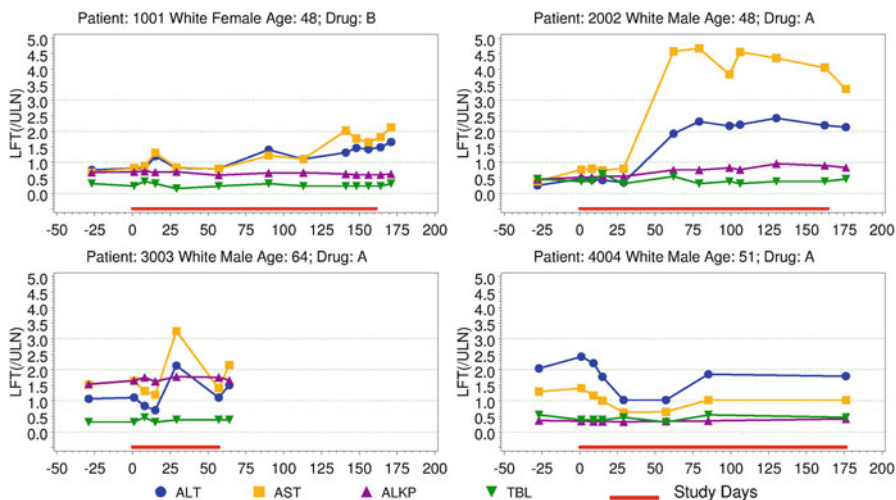
Fig. 17.8 Customized patient profile display of LFT measurements

17.3.5   *Patient Profile of LFT Measurements*

As previously noted, much of safety analysis is concerned with individual subjects rather than summary data. Figure 17.8 provides a powerful and concise summary of liver function information on selected individual subjects (in this case, on the basis of any LFT exceeding 2xULN during the trial).

All 4 liver function parameters are plotted as a function of time for each individual subject, allowing ready assimilation of several pieces of information. These include the time course of the elevations relative to treatment, the presence or absence of simultaneous elevations, outcomes of dose interruptions, dose reductions as they relate to the elevations and outcomes of a patient subsequent to an interruption or reduction. When data from many individual patients are needed, a series of displays of this kind can be produced. Once the first screen has been viewed and understood, the remainder can be quickly assessed as long as the display style is kept consistent.

This graph was drawn using SAS™.

17.3.6   *Other Possibilities*

Another aspect of laboratory data that can be of great interest is the way in which AESIs accumulate over time under different treatment regimes. This can be effectively displayed with a cumulative incidence plot, as in Fig. 17.2.

Table 17.2 Other clinically meaningful lab measurements to consider

Category	Clinical interpretation	Measurements
Haematotoxicity	Grades 3 and 4	Red cell count, mean cell volume, platelets, haematocrit, haemoglobin, reticulocyte count, white cell total count and differentials
Nephrotoxicity	Creatinine increase from baseline >0.3 mg/dl (as defined by Acute Kidney Injury Network)	Serum creatinine, blood urea nitrogen and creatinine kinase
Lipids		Cholesterol, thyroxin, LDL, HDL and triglycerides
Rhabdomyolysis/ muscle injury	By profiling	Creatinine phosphokinase (CPK), AST, lactate dehydrogenase (LDH), red cells or myoglobin on urine test
Paediatrics	Growth; CNS functioning; reproductive or endocrine status	Height, weight, BMI, often transformed into z-scores; serial assessments of IQ, as an example, for long-term studies; androgens, estrogens and relevant hypothalamic hormones
Suicidality	Incidence rate (e.g., in a forest plot)	Ideation, attempts, deaths

There are many other lab measurements that could readily be displayed in graphical form. Table 17.2 shows a list of lab measurements to consider. All of the graphical methods described above may be applied to these other lab variables.

17.4   Cardiac Safety

One of the major issues that have led to drug withdrawals has been cardiovascular incidents, so there was an early focus on cardiac safety in safety data analysis. QT prolongation and Torsades de Pointes is of primary concern, but any conduction-interval prolongation (e.g., PR prolongation) could be a potential safety concern and possible showstopper for a new drug. Figure 17.9 shows a stylized ECG trace annotated with the letters from which some of the heart rhythm measurements are associated. Other crucial issues are heart failure (predicted by ejection fraction) and myocardial infarction (predicted by troponin levels).

The main derived endpoints from heart traces are the RR, PR, QRS and QT intervals. The last of these, usually in a corrected form and called the QTc interval, is the key one for general studies. It is a marker for cardiac toxicity: prolonged QTc can lead to increased risk of Torsades de Pointes (TdP)—a rare but life-threatening arrhythmia. This is a Sentinel Event that the FDA require reporting as soon as there is awareness of a case. FDA guidance on evaluation of QTc is as follows:

- Increases to >500 ms are of clinical concern
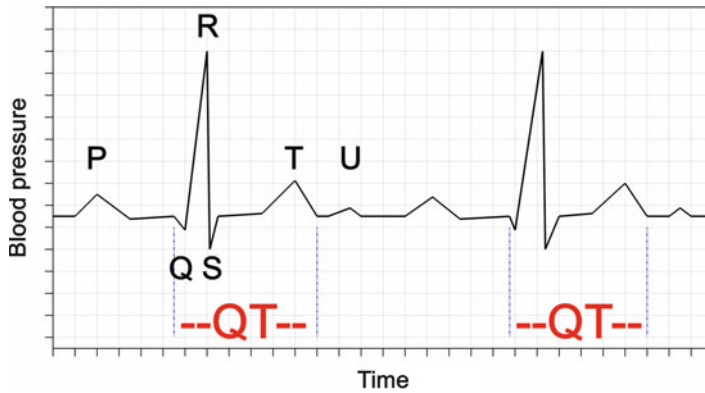- Increases to >480 and 450 ms are also of interest

Fig. 17.9 A stylized ECG trace

- Changes from baseline of >60 ms (regardless of absolute value) are of serious clinical concern
- Change from baseline of >30 ms are of clinical concern

Relevant questions to ask based on the guidance are:

- Is there a significant change over time in the distribution of QTc results?
- How many people report a significant shift in QTc values, i.e., an increase of >30 or >60 ms?
- How many subjects report a QTc interval of >450, >480 or >500 ms?

Many of the graphical displays in the previous section are clearly appropriate to cardiac measurements like QTc, such as the scatterplot of shifts in Fig. 17.4.

17.4.1 Cumulative Distribution Plot of QTc

For a comparative trial, a plot showing the detailed distributions of critical measurements like QTc can be invaluable for giving reassurance or highlighting areas of concern. Figure 17.10 shows a cumulative distribution function (CDF) plot from the Safety trial.

This display allows close scrutiny of the distributions, with reference to clinical criteria. Note that the percentage of subjects with a change greater than 0 is just over 10%. This is as expected, as explained before: QTc was measured at baseline and at 8 visits, so the chance that the baseline measure is the smallest of these would be 11% (i.e., 1/9) if the drugs have no effect and successive measurements on a patient
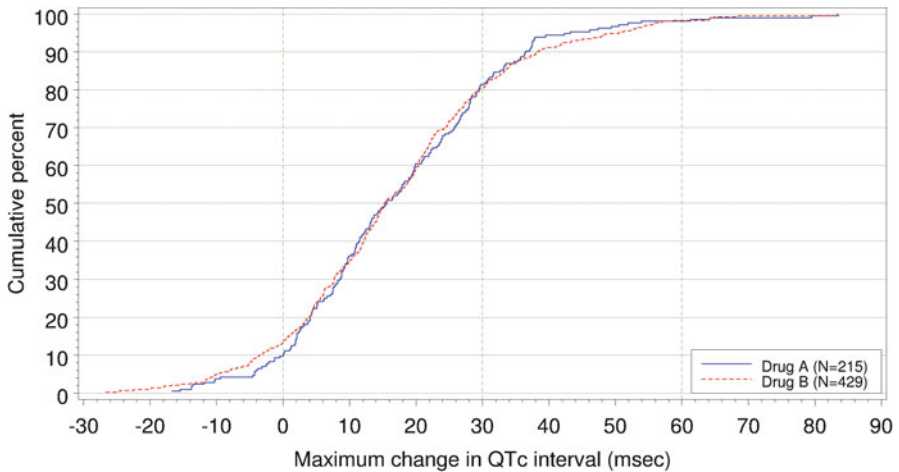
Fig. 17.10 CDF plot of maximum QTc changes

can be taken as independent. Note that it is particularly useful to show a faint grid with this graph, as this helps detailed interpretation of any differences noted between the two step functions.

This graph was drawn using SAS.

17.4.2 Boxplot Profile of QTc

The boxplot, illustrated in Fig. 17.7, can also be used to display the change in the distribution of a variable over time profile. Figure 17.11 compares the distribution of QTc changes from baseline on the 2 treatments arms of the Safety trial, for each of the 8 visits. In addition, a right-hand margin has been added to show the distribution of the maximum change from baseline. The numbers of patients measured at each visit is also displayed in a bottom margin, as before, but here the margin has been brought inside the frame, which can help to emphasize the relationship between the values and the plotted information.

Each individual point representing a change greater than 60 ms has been labelled here with the patient number: this draws attention here to Patient 194, who had an increased level of QTc from Week 12 onwards. Some labels are overwritten, but it would be difficult to arrange to separate them; clearly, the amount of labelling needs to judged carefully if it is to be of use. An alternative to labelling would be to list the values of concern in a separate table.
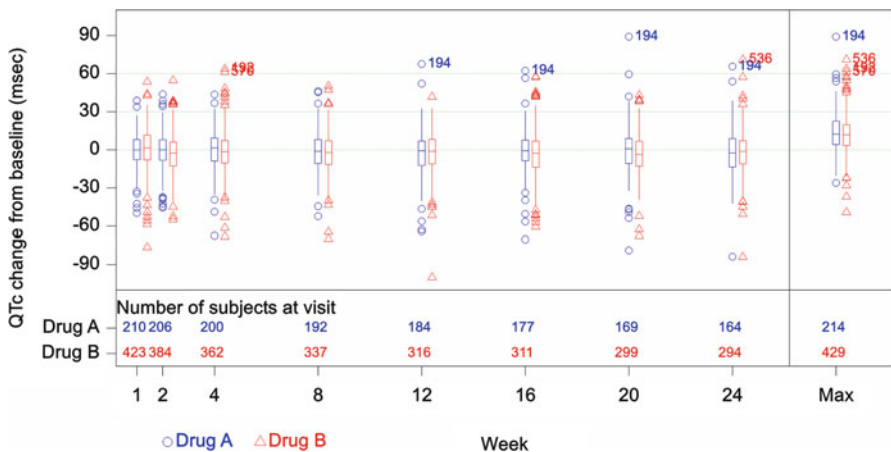
This graph was drawn using GenStat™.

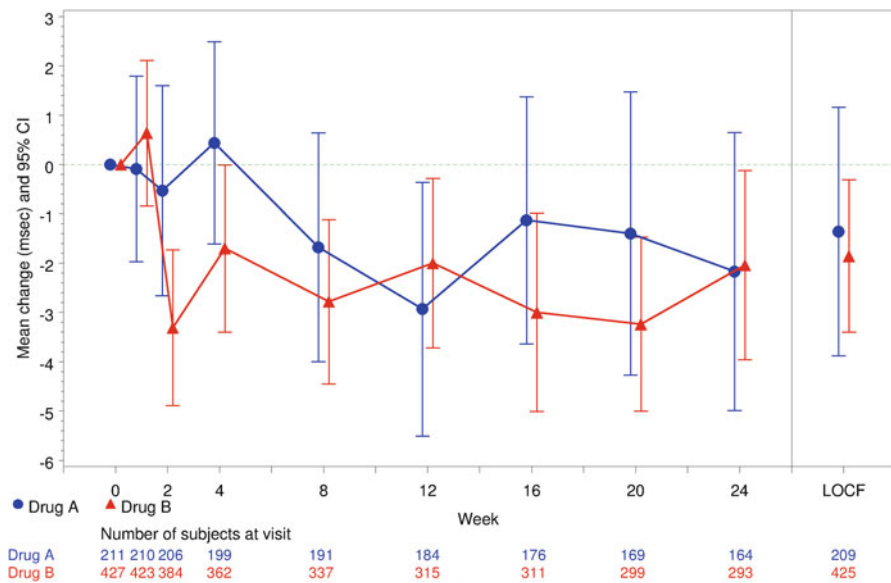Fig. 17.11   Boxplot profile of QTc changes from baseline

Fig. 17.12   Mean change from baseline (and 95% CI) in QTc over time

17.4.3   Mean Profile of QTc

As well as showing the individual subject data in the above distribution plots, it may
be useful to focus on the evidence for systematic difference between the drugs.
Figure 17.12 shows the mean QTc changes over time, with confidence limits to put
the small differences into context.

This kind of display is commonly seen for reporting efficacy, showing how the effect of treatments compared in a trial change over the course of the trial. Usually, it is the difference at the end of the trial that is of primary interest, and that difference may be adjusted to try to take account of interfering factors, such as drop-out of patients from the trial. One such method is "last observation carried forward," as shown here, but more advanced methods using multiple imputation or mixed modelling are now preferred (Mallinckrodt et al. 2008).

This graph was drawn using SAS™.

17.5 Conclusion

We have described 11 different graphical designs that we recommend for displaying safety information. We used each design to display particular safety outcomes, but many of the designs can of course be used for a wide range of different outcomes, and indeed for efficacy outcomes as well. We have omitted a large class of graphical designs, which are being used increasingly in pharmaceutical companies to monitor safety of drug development programmes. These are interactive designs which allow the viewer to modify the display using a graphical interface or drill down to find further information about aspects of interest in the initial display. Other chapters in this book describe these, in particular Chap. 10.

References

Amit O, Heiberger RM, Lane PW (2008) Graphical approaches to the analysis of safety data from clinical trials. Pharm Stat 7:20–35

Andrade RJ, Lucena MI, Fernández MC, Pelaez G, Pachkoria K, García-Ruiz E, García-Muñoz B, González-Grande R, Pizarro A, Durán JA, Jiménez M, Rodrigo L, Romero-Gomez M, Navarrao JM, Planas R, Costa J, Borras A, Soler A, Salmerón J, Martin-Vivaldi R (2005) Drug-induced liver injury: an analysis of 461 incidences submitted to the Spanish registry over a 10-year period. Gastroenterology 129:512–521

Cleveland WS (1993) Visualizing data. Hobart, Summit, NJ

Crowe BJ, Xia HA, Berlin JA, Watson DJ, Shi H, Lin SL, Kuebler J, Schriver RC, Santanello NC, Rochester G, Porter JB, Oster M, Mehrotra DV, Li Z, King EC, Harpur ES, Hall DB (2009) Recommendations for safety planning, data collection, evaluation and reporting during drug, biologic and vaccine development: a report of the safety planning, evaluation, and reporting team. Clin Trials 6:430–440

Lammert C, Einarsson S, Saha C, Niklasson A, Bjornsson E, Chalasani N (2008) Relationship between daily dose of oral medications and idiosyncratic drug-induced liver injury: search for signals. Hepatology 47:2003–2009. doi:10.1002/hep. 22272

Mallinckrodt CH, Lane PW, Schnell D, Peng Y, Mancuso JP (2008) Recommendations for the primary analysis of continuous endpoints in longitudinal clinical trials. Drug Inf J 42:303–319

Pintilie M (2007) Analysing and interpreting competing risk data. Stat Med 26:1360–1367

Pocock SJ, Clayton TC, Altman DG (2002) Survival plots of time-to-event outcomes in clinical trials: good practice and pitfalls. Lancet 359:1686–1689

Tukey JW (1977) Exploratory data analysis. Addison-Wesley, New York, NY

Watkins PB (2005) Idiosyncratic liver injury: challenges and approaches. Toxicol Pathol 33:1–5

Wilke RA, Lin DW, Roden DM, Watkins PB, Flockhart D, Zineh I, Giacomini KM, Krauss RM (2007) Identifying genetic risk factors for serious adverse drug reactions: current progress and challenges. Nat Rev Drug Discov 6:904–916