Drug Safety Reporting- now and then

David J. Garbutt, Business & Decision, Zürich, Switzerland

ABSTRACT

INTRODUCTION

This paper is about the now and then of safety reporting, about its future and where it can, and should, go. I will not talk about Drug Safety Monitoring, although many of the lessons I hope to convince you about could apply there also.

Many of you may know the story of the emperor who had no clothes on, although he had been convinced he really did. Here we have a big pile of clothes, but no emperor. Our journey today is to see how we can go about putting the emperor back into his clothes so he can easily be recognized, as an emperor that is.

This paper will remind us why we do Safety Reporting, and ask if what we currently produce really fills that need, what we could do to improve our product, and briefly look at factors that I believe that indicate safety reporting will change in the next few years.

CLOTHES, BUT NO EMPEROR

Standard Safety reporting generates large amounts of paper. Listings with 20,000 lines are not uncommon. And AE tables can be as big, not to mention shift tables. A colleague of mine recently had to review a shift table taking up 280+ pages; actually there were four tables that long. Is there a dummies guide to interpreting shift tables? I certainly hope there is a dummies guide to programming them [1].

This sheer amount of product creates problems in generation, assessment, validation, assembly and last, and worst – comprehension and communication. Safety outputs are almost always descriptive – the outputs we create are only rarely analytical and therefore very limited. And, I have always suspected, not read.

We aim to show a drug has no dangers, or at least we can make clear what dangers there are, and under what circumstances they are important to the patient. We should also be asking what constellation of AEs comes with the drug. Is the incidence dose or exposure related? Is it related to any concomitant medications? Are there any particular-prone patient subsets? Are there any surprises in the data?

SAFETY DATA ARE MORE IMPORTANT THAN EVER

Safety reporting used to be a check but now it is vital to marketing, drug screening, approval, and perhaps continued existence on the market.

Good safety analysis also has the potential to affect time to market. A 2003 study at the FDA[§] of the reasons for repeated reviews of new drug applications (NDAs) showed the commonest reason was safety concerns.

> Standard NMEs studied were those with total approval times greater than 12 months in 2000 and 2001. Fifty-seven percent of these applications had times greater than 12 months, ranging from 12.1 to 54.4 months. The most frequent primary reasons for delay on the first cycle were **safety issues (38 percent)** followed by efficacy issues (21 percent), manufacturing facility issues (14 percent), labeling issues (14 percent), chemistry, manufacturing, and controls issues (10 percent), and submission quality (3 percent).
>
> Source: http://www.fda.gov/cder/pike/JanFeb2003.htm

For priority NDAs the proportion of delays due to safety was 27% and came second to manufacturing and quality concerns.

[§] FDA refers to the Federal Food and Drug Administration of the US Government. Not to be confused with Functional Data Analysis mentioned later.

CHARACTERIZING SAFETY DATA:

Safety data are not easy data to analyse with conventional statistical methods because many of the 'standard' assumptions are not fulfilled. Pathological features frequently seen in safety data include:

Asymmetric non-normal distributions, often with variability proportional to mean, heterogeneous subpopulations (e.g. patients are differentially prone to AEs – for example liver damage). Data are available as counts or time to occurrence. There are large amounts of variability –e.g. clearly seen at baseline. The data are inherently multivariate time series. There are scaling and range shifts between centres. Differentially responding subgroups of patients may have varying frequencies across centres.

Adverse events

Count data for Adverse Events is variable (as count data is) and complicated by the large number of possible events, and the high incidence on placebo. The large number of possible events means there is a great example of the possibility of false positives because so many tests are being performed. Some methods of analysis break down because there are many zero counts on placebo treatment.

ECG Data

These data are increasingly collected (especially in phase II) and are multivariate, non-normal, longitudinal series of measures per patient. And in fact the measurements are summaries derived from from traces measured at two or three time points. The derivation of these measures needs a certain skill and this introduces another source of variation. In addition the assessment of abnormalities is not very reproducible between experts (20% of cases will be assessed differently).

Laboratory test result data

They have some similarities to ECG data – they are also multivariate, non-normal, correlated time series per patient. They are typically assessed using codings comparing the values to (more or less arbitrary) normal ranges. These limits are a univariate approach which is well known from basic multivariate distribution theory to be problematical for correlated variables [2]. For an example see Figure 1 due to Merz [3]. This figure shows how high the misclassifications can be using this method. And these misclassifications go both ways – signals missed that should not have been (FN in figure) and *vice versa* (FP).

Against lab normal ranges

Normally we accept normal ranges at face value and I have always wondered how they were derived. (One reason is that we have skewed data and estimating quantiles (like 95% for example) needs a lot of data to be accurate. Ignoring the skewed shape and using theoretical limits based on a normal distribution would be misleading. A 1998 paper assessing lab normal ranges against a large (8000+ people) population found a situation of concern.

> Abstract:
>
> BACKGROUND: When interpreting the results of clinical chemistry tests, physicians rely heavily on the reference intervals provided by the laboratory. It is assumed that these reference intervals are calculated from the results of tests done on healthy individuals, and, except when noted, apply to people of both genders and any age, race, or body build. While analyzing data from a large screening project, we had reason to question these assumptions.
>
> METHODS: **The results of 20 serum chemistry tests performed on 8818 members of a state health insurance plan were analyzed.** Subgroups were defined according to age, race, sex, and body mass index. A very healthy subgroup (n = 270) was also defined using a written questionnaire and the Duke Health Profile. Reference intervals for the results of each test calculated from the entire group and each subgroup were compared with those recommended by the laboratory that performed the tests and with each other. Telephone calls were made to four different clinical laboratories to determine how reference intervals are set, and standard recommendations and the relevant literature were reviewed.

armin 8/10/08 11:06

Comment [1]: The monitored electrical activity of the heart is a time series, but not multivariate. Normally, this time-series is transferred into summary measures, which are strongly related and therefore correlated. I do not see to have multivariate, correlated time-series. Either you have time series, or multivariate, correlated numbers (summary measures extracted from the time series), but not both, isn't it. Most likely you are referring to repeated monitoring over time. Then we have the multivariate, correlated summary measures repeatedly over time. I would in this case not speak of a time series, as time-series are often considered as such only when having thousands of time points, in contrast to longitudinal data, panel data or repeated measures data.
What I want to say is: I did not quite follow to what you refer to. Possibly, it is crystal clear for the targeted audience and does not need more explanations.

RESULTS: **The results from our study population differed significantly from laboratory recommendations on 29 of the 39 reference limits examined,** at least seven of which appeared to be clinically important. In the subpopulation comparisons, "healthy" compared with everyone else, old (> or = 75 years) compared with young, high (> or = 27.1) compared with low body mass index (BMI), and white compared with nonwhite, 2, 11, 10, and 0 limits differed, respectively. **None of the contacted laboratories were following published recommendations for setting reference intervals for clinical chemistries.** The methods used by the laboratories included acceptance of the intervals recommended by manufacturers of test equipment, analyses of all test results from the laboratory over time, and testing of employee volunteers.

CONCLUSIONS: Physicians should recognize when interpreting serum chemistry test results that the reference intervals provided may not have been determined properly. Clinical laboratories should more closely follow standard guidelines when setting reference intervals and provide more information to physicians regarding the population used to set them. Efforts should be made to provide appropriate intervals for patients of different body mass index and age.

Mold JW, Aspy CB, Blick KE, Lawler FH (1998) [4]

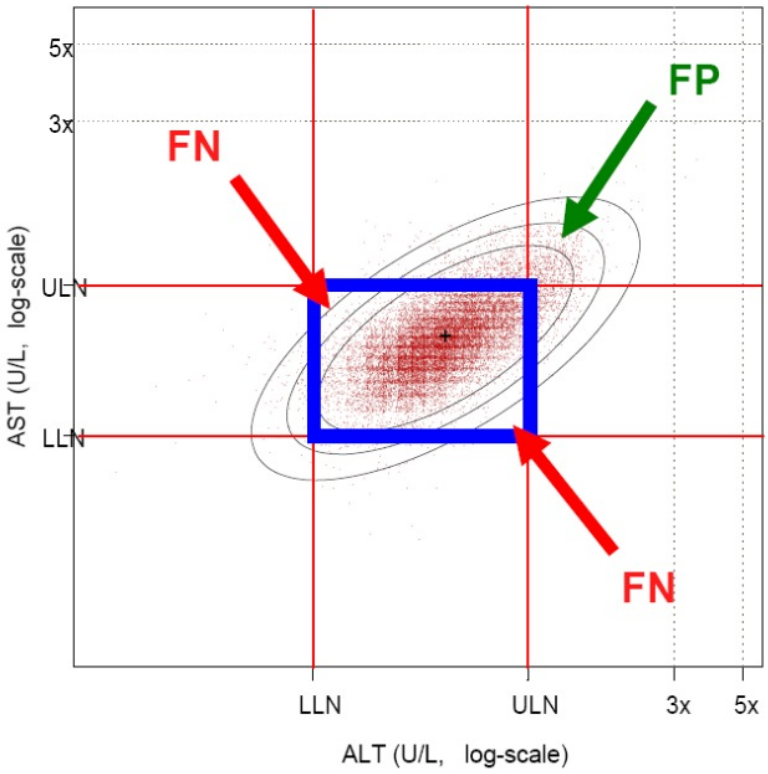


Figure 1 Univariate limits are misleading for correlated variables. FN is a false negative, and FP a false positive. Figure from Merz [3]

The situation may have improved now, although a recent survey of 169 laboratories by Friedberg et al, (2007) [5] would seem to argue that things have not changed. In any case this is just another argument for using the internal properties of the data we have rather than discarding information and using arbitrary classifiers.

Levels of variation in safety data

There are multiple sources of variability in safety data and this must be taken into account when analyzing and preferably plotting. There are large differences between patients and with many repeated measures there are visit to visit correlations. The time scale of these correlations varies according to the lab parameter being observed – but for blood haemoglobin levels it should be at least 3 months (this being the replacement time for blood haemoglobin). So simple scatter plots of individual liver function enzymes (Figure 15) ignoring the patient dimension are liable to be misleading. It is also a corollary that repeated measurements are not worth as much might be expected. On the plus side we are lucky to have baseline data for treated patients and placebo patients during the whole time course of treatment. In cross-over trials we can estimate treatments differences within patients and escape even more variation.

WHY IS THERE NO MORE ANALYSIS THAN THIS?

> Unlike those for efficacy endpoints, clinical hypotheses for safety endpoints are typically loosely defined. This often results in little attention being given to additional and more innovative approaches (including graphical ones). In a recent informal survey among over 200 statisticians involved in clinical trial design and analysis in GlaxoSmithKline, fewer than 5% routinely used graphical approaches to summarize safety data.
>
> Amit, Heiberger, & Lane (2007)[6]

I find this state of affairs shocking, although it fits with my experience of what reporting is done currently and what has been standard practice for the last 20 years.

I suspect the number using any (statistical) analytical method is even lower. And consider for a second how much money is spent on making lab tests on patients. We are looking at hundreds of dollars per time point, per replication. With a single patient's lab data costing thousands of dollars - we should ask how much programming time would that money buy? And how much reading and reviewing time it might save?

A new paradigm for analysing of Safety Data

It may be worth going so far as to say that the analysis of safety data should be aimed at identifying patients with unusual patterns of response and characterizing exactly what those responses are.

WHAT CAN WE DO?

It is difficult to prove a negative - that there are no dangers. Because there are many rare effects – such as *Torsade des Pointes* – with an incidence of 1 in 100,000 in the general population.

If our drug increases the chance of that condition by a factor of 10, we still need to study thousands of patients to have a reasonable chance of detecting the problem. It is all dependent on the power of the test - How many patients? How long (in total) have they been exposed to the drug?

With Safety data we really want to prove the null hypothesis - but not fall into the trap of declaring an issue when there is not one. So we comprehensively look for issues - but not analytically. So, too much is left to ad hoc comparisons, which is not better. We group values and lose information (e.g. lab shift tables). We do simplistic univariate analyses. We list or tabulate endless subsets, without proper comparison.

We have a problem because more data are coming. How can we include genetic markers in this informal mess?

Undersized and over-clad

Efficacy analysis has always been more important and because of this studies are sized for tests planned for efficacy variables and undersized for accurately measuring safety issues. I believe another reason is that safety data are more amenable to standardisation and in many companies this was done 10-15 years ago according to good (or acceptable) practices at the time. Standardisation is good and saves money and needlessly repeated effort, but setting things in stone is also like fossilisation.

TEN YEARS AGO IN COMPUTING:

***Intel** released the **333 MHz Pentium II** processor with MMX instructions and a 66 MHz bus. It incorporated a 0.25 micron CMOS manufacturing process. (This is roughly 1000 times larger than today's consumer chips).*

***April 20 1998**- at a public demonstration of Windows 98, Bill Gates crashes the operating system.*

Apple unveils a **15" monitor**

(Source: http://www.islandnet.com/~kpolsson/comphist/comp1998.htm)

SAS 6.11 is the current version

SAS/**Graph** was 14 years old

WHY IS TEN YEARS AGO DIFFERENT?

Computer resources and software, especially, were different then, and the methods for creating graphics output for publication had a much longer turn around time than now. Although we thought (and we were right!) that a week was a big improvement on waiting for the time of a technician to make a black ink drawing with a Rotring® pen and tracing paper.

Statistics and statistical software have not stood still in the last 15 years. There are pictures, models and software that can help us.

MAKING PROGRESS

Modern statistical graphics was started by Tukey with his book Exploratory Data Analysis (EDA, published 31 years ago in 1977)[7]. In this book he developed and exemplified methods for examining data which were semi-analytical. By this I mean they were graphical but were also based on an underlying method. A good example of this is his treatment of two-way tables. He also developed the boxplot for displaying the distribution of data while exposing asymmetry and the presence of outliers. EDA is full of hand drawn graphs; at that time sketching was the only way to quickly examine a set of data points. This exposes an important aspect of systems. The effort of making a plot 'for a quick look' should be low enough to make speculation no effort. And when something interesting is found the effort to create a report quality output should also be as low as possible.

The development of statistical graphics really took off in the 80's and 90's with the work of Cleveland [8], Tufte [9] and others which utilised experimental work on our perceptual mechanisms and a realization that good communication would result from designs made with those characteristics in mind.

That research and the rise of interactive statistical packages has made these methods mainstream. There have been good implementations available for some time in S-Plus, R, and JMP.

NEW GRAPHICS OF NOTE

The advent of lattice and trellis graphics and high resolution displays really made the use of static plots a viable method of data analysis. It is an important development because not all data is analysed statistically or by statisticians, much data analysis is done by scientists. Producing sets of plots

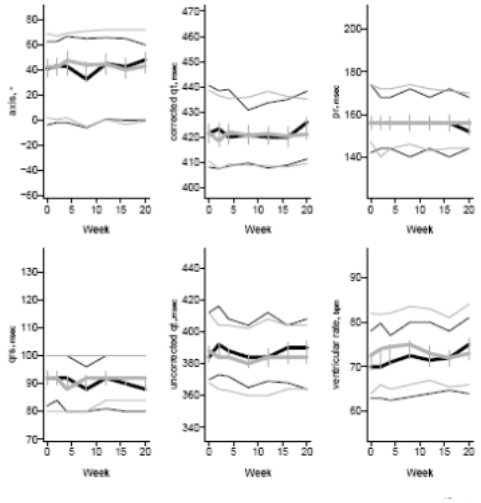


Figure 2 Quartiles of EKG variables in two treatments over time (Harrell, [12])

armin 8/10/08 08:34

Comment [2]: I do not like this figure having two captions. I would remove the original one and move its text to the one below. Especially the start "Figure 10" is confusing.

conditioned on other variables can really show what factors are important and they are especially useful when analysing data sets where the factors used for analysis have interactions. I have mentioned several books for those wanting to read more on this subject I should also mention Frank Harrell's graphics course which is available on line [10]. A useful survey with history and examples is Leland Wilkinson's article on Presentation Graphics[11].

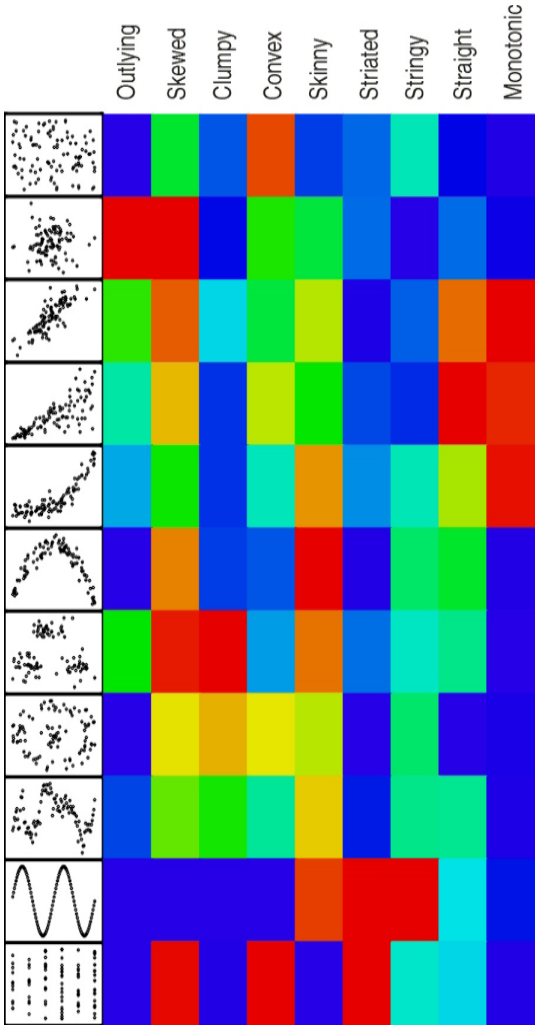
I will illustrate some of the new methods later in this paper but for now I will just mention some of the

most useful. Dotplots should replace barcharts as they are easier to interpret, more flexile and use less ink. Sparklines (Tufte [9]) put graphics in-line as word sized pictures. Like this  or this . The aim is to integrate argument and evidence in text, but sparklines have also been used to achieve high information densities, a web page of examples is maintained at http://sparkline.wikispaces.com/Examples. There are obvious possibilities here for plotting data for large numbers of patients on a single page for use as a compact patient profile.

There is some very interesting work by Frank Harrell in Rreport [12] which re-imagines the independent safety monitoring board (DSMB) report and displays of ECG data using half confidence intervals (see Figure 2) as well as time to event analyses for adverse events. The plots in Figure 2 also use shades of grey in a clever way by plotting the treatment in grey after the placebo (in black) so differences are highlighted. The outer lines are 25% and 75% quantiles and the thicker bars depict medians. Vertical bars indicate half-widths of approximate 0.95 confidence intervals for differences in medians. When the distance between two medians exceeds the length of the bar, differences are significant at approximately the 0.05 level. The comparisons of demographic data across treatments groups shown in the sample report is also interesting.

The HH R package [13] which accompanies the book by Heiberger and Holland [14] includes the R function Ae.dotplot. We will see examples and adaptations of this later.

Figure 3 Scaled graph-theoretic measures (Wilkinson *et al*. [17])



Another approach which is not strictly graphical and not strictly modelling is scagnostics from John Tukey but never fully developed by him [16]. The term is a(nother) neologism coined by John Tukey and is a portmanteau word derived from **sca**tter plot dia**gnostics**. The aim is to quantitatively classify a two way scatter plot and therefore allow an automated or semi-automated search for the interesting features in data. This looks like a promising approach for cases where we look for 'issues' without being able to specify in advance all the possible patterns that might be important. Wilkinson *et al.* [17], [18] have expanded and extended this work.

Although there is no direct model the measure is quantitative and therefore open to simulation and statistical calibration. There is a scagnostics R package [19]. As an illustration consider Figure 3. This shows the results for nine (scagnostics) measures (labelled outlying, skewed, clumpy, etc.) with a heat map (blue is low and red high) of their value as calculated on eleven example scatter plots that are shown in the leftmost column. The way that properties of the plots are captured by the measures is almost uncanny. One reason I believe this approach may be useful is that we need to take account of the multiple levels of variation in our data. Two of the main ones are patients/subjects and visits (~time on drug). Surveying scatter plots of variables per patient and finding the few that show a response looks like a good strategy.

New graphics are not just available for quantitative data. The invention of tree maps in the early 90's provided a flexible way to visualize qualitative variables more normally shown with contingency tables and modelled with log-linear models. A good introduction to the tree map is at http://eagereyes.org/Techniques/Treemaps.html they are available to SAS users via macros written by Michael Friendly [15]. They are available as standard in JMP.

Plotting larger amounts of data

Plotting quantiles as in Figure 2 is a marked improvement over means and standard errors (since we have little faith the distributions are normal with homogenous variances), but one of the lessons from Cleveland's work [8] is to plot all the data. This works very well for "small" amounts of data but when the number of points is in the hundreds over-plotting starts to hide information. In our kinds of data clustering may indicate there are patient subgroups that are responding differently. In these cases it makes sense to plot the data density (rather than individual points) along with the smoothed or fitted values. One way to do this is hexagonal binning which is very fast even for very large numbers of observation. The R package hexbin is available [20]. Another possibility is the two dimensional HDR (High density region) boxplot function from the hdrcde R package by Hyndman [21] which gives very pleasing results. Figure 4 shows two such plots with fitted lines and density estimates for SAS 9 migration times. There are about 400 studies measured and these plots show total migration time vs. total size of the SAS data in each study.

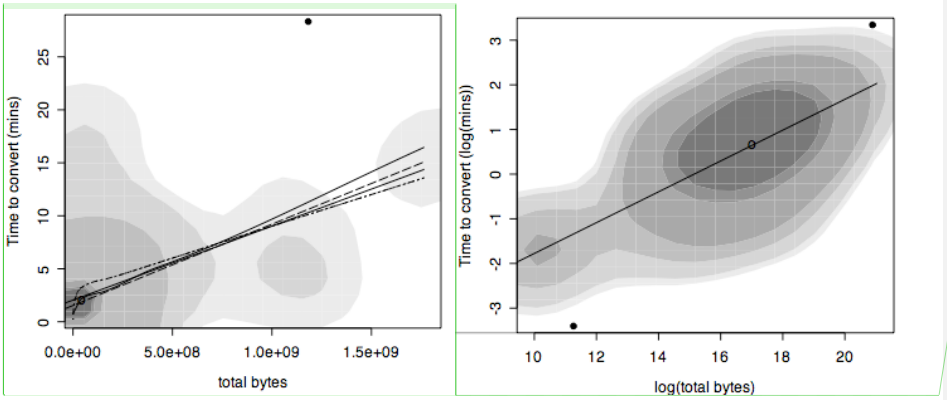


Figure 4 Plot of Time to convert vs.Total Bytes showing smoother and HDR-2d boxplot of points. The left hand panel shows raw data and the right shows log transformed data. Contours are drawn at probabilities of 1%, 5%, 10%, 25%, 50% & 60%. Note one point is outside the 1% boundary on both plots, and one is only visible as an outlier on the right.

R Program to create Figure 4.

```
# program by DJ Garbutt 27.Oct.2007
# load the library - assume it is already installed
```

armin 8/10/08 08:42

Comment [3]: There should not be any data below 0 on any of the two dimensions. This is a matter of correctness (it looks slightly misleading know), but since this should be clear enough, you do not necessarily need to correct this.

armin 8/10/08 08:42

Comment [4]: I see two points outside the 1% boundary in the right panel.

```
library("hdrcde")
# get the data read in on a previous session and saved
attach(prf)
#--------------------- left panel
# make the 2-d boxplot
hdr.boxplot.2d( X..SAS.bytes,Total.Seconds/60,
   xlab="total bytes",
   ylab="Time to convert (mins)",
   prob=c(0.01,0.05,0.10, 0.25, 0.5, 0.6)
  )

# add smoothers of varying 'smoothness'
lines(loess.smooth(X..SAS.bytes,Total.Seconds/60))
lines(supsmu(X..SAS.bytes,Total.Seconds/60, bass=2, lty=3))
lines(supsmu(X..SAS.bytes,Total.Seconds/60, bass=2), lty=3)
lines(supsmu(X..SAS.bytes,Total.Seconds/60, bass=1.2), lty=4)
# Fit robust line to data
rlm.prf <- rlm(Total.Seconds/60 ~ X..SAS.bytes, data=prf)
# add fitted line to plot
abline(rlm.prf, lty=5)
# save plot in two formats (Mac OS X)
quartz.save("timeVSbytes.png",type="png")
quartz.save("timeVSbytes.pdf",type="pdf")
#---------------- right panel
# redo with log transform on both axes
hdr.boxplot.2d( log(X..SAS.bytes + 1),log(Total.Seconds/60),
   xlab="log(total bytes)",
   ylab="Time to convert (log(mins))",
   prob=c(0.01,0.05,0.10, 0.25, 0.5, 0.6)
   )
# make  robust linear fit
rlm.prf <- rlm(log(Total.Seconds/60) ~ log(X..SAS.bytes + 1), data=prf)
# add the line from the fit to the plot
abline(rlm.prf)
# save picture as png (MacOS X )
quartz.save("logtimeVSlogbytes.png",type="png")
# NB panels created separately and juxtaposed in word. Could better easily
# make a two frame plot in R
```

An improvement of this plot would be to avoid plotting high density regions below zero, since for our data zero is a natural minimum.

Analysing more data

Typically (perhaps, universally) safety data are assessed on one trial with, usually after approval, safety updates. But pharma companies are actually sitting on large amounts of placebo and baseline data. These data provide an enormous – but unexploited pool of data to calibrate safety analyses. It is tempting to think that placebo is the same treatment for all drugs – and this is so, but there is a subtle trap here – the selection of patient populations. Depending on the drug and trial patients might be very ill (perhaps with terminal cancer, or late stages of heart disease) or relatively healthy – perhaps in a trial of cold vaccine. However recent advances in clustering and classification mean that this approach might be viable when combined with patient matching.

Analysing all the data and keeping the patient as unit

Given the under-powering of trials for safety purposes it would make sense to restrict safety analyses to a single document covering all trials done so far, i.e. what is sometimes called an integrated summary of safety (ISS) and keep a bare minimum of analyses in the individual studies reports. There would be several

armin 8/10/08 08:54

Comment [5]: Study participants know in what kind of investigation they are taking part and what the active drug, they possibly receive, roughly does. This alters there perception and also how they may react to a placebo treatment. Patients in completely different studies with the same placebo treatment may show different reactions solely based on the different perceptions they have, as a consequence of seeing the whole treatment in the context of the study background.
Secondly, the administering scheme of the placebo may also alter the perception on what it is or does and the outcome, as a consequence. Comparing placebo treatments between different studies with different administration schemes, dosages or form of drug may therefore not be possible.

obvious advantages for this approach. The ISS could be updated after every trial and consistency is made easier when there is one definitive analysis. It also means that small signals just below the bar in individual trials would not escape attention so easily. It would also be advantageous because as new ideas or results come forward we re-analyse all data by default rather than deciding if old trial reports should be re-opened. It would also be possible to keep this concept as new indications are added. More patients for analysis means more power for detecting real issues and less chance of extreme results from one trial. This approach carries forward nicely into phase IV and perhaps even into safety monitoring.
It could even be argued that the unit of the safety analysis should always be the patient and the dose and trial are just blocking variables. This approach could be feasible given the advances in meta-analysis and mixed models and the data standardisation from CDISC.

NEW STATISTICAL MODELS

Many new models have been created and fitted to data in the last 30 years and a few examples with relevance to safety data (and which tackle the statistical issues mentioned above) are:

- Quantile regression
  Fit arbitrary curves without the need for a guessable functional form.
- Data smoothers like loess and supsumu [22, p231]
- Bayesian analysis of AEs by Berry & Berry (2004) [23]
  Hierarchical model - AEs within patients, within body systems addressing the multiple comparisons problem and using information from other AEs in the same body system. Available in Insghtful's Clinical Graphics application (iCG.
- Bayesian analysis of Liver Function tests [24]
- Multivariate analysis of Liver enzymes data [25]
- GAMMs [22, p232]
- DHGLMs (model variance as well as means)
- Functional data analysis
  [see http://www.psych.mcgill.ca/misc/fda/ for intro]
- Perfect (and cross-validated) subset analyses with partition tree models [22, ch 9]
- Model-based clustering
- Meta-analysis
- State-space models and non-Gaussian processes in time series analysis

Partition tree models are worth discussing further because they are an analytical way of finding subsets of patients that have effects and could also be used in cases where we wish to show that patient selections are the same. They are rigorous because they can be cross-validated and specifiable in advance to be used. I am not aware of anyone using them in such contexts.

armin 8/10/08 09:01

Comment [6]: Do you mean "Insightful Clinical Graphics". I did not find other useful references to iCG and the Insightfuly Clinical Graphics was not referred to as iCG. Are you sure that the Berry&Berry model is fully implemented in this tool (and it does not (via some complicated interface) call BUGS to do the job? I did not find any more information on this over the web.

NO TABLES, MORE GRAPHICS

This mantra has become a movement even for statistics journals with the publication of the paper Practice what we Preach? by Gelman, Pasavia & Doghia [26] which takes an issue of a statistics journal and develops graphical displays that improve on all the tables they contain. Another paper 'Tables to Graphs' by Kastellac & Leoni (2007) has an accompanying web site as well [27]. This point of view is also shared by at least some at the FDA see recent talks by Bob O'Neill and Matt Soukup (2007)[28]. Other recent advocates of graphics are Wang (2007)[29] , O'Connell (2006)[30], Merz (2006)[3].

WHY CAN IT IMPROVE NOW?

GRAPHICAL METHODS ARE BECOMING IMPORTANT AT THE FDA

Within the FDA the advent of the JANUS data warehouse system means that statistics reviewers are moving to a situation where they will have easy access to all data from all submissions and be able to re-analyse the data themselves. There are several advocates of graphical analyses there so it is fair to assume 'the FDA will reanalyze your data this way'. It is therefore prudent to use the same techniques and find the insights they bring first, if only to be better prepared when answering questions. Because they will have the data available in a standardised form they will be better able to develop graphical analyses.

BETTER GRAPHICS SUPPORT, NEW SYSTEMS ARE AVAILABLE NOW

There is a growth of packaged solutions becoming available notably iCG from Insightful. There is also the PPD graphical patient profiler and tools from Phase Forward, and Spotfire.

Roll-your-own solutions can choose from many systems most notably JMP (from SAS) and R (perhaps with ggplot and ggobi).

Coming 'soon' is SAS 9.2 and the new graphics procedures using templates. Search for sgplot and other sgxxx procedures on the support.sas.com website. This looks a promising option because of the templating that would allow a high level of re-use than is possible with current SAS/Graph plots. However the crucial issue will be how generic the templates can be – can they, for example, take plot labels automatically from a variable's label?

COSTS

Why produce unneeded paper output? FDA has stated that for submissions planned with SDTM the amount of listing to be provided is 'negotiable'. It wouldn't make sense to deliver more in this case, so it can only mean a pressure towards less paper and perhaps more analytics.

CDISC IS COMING AND CREATING A NEW SOFTWARE MARKET

The advent of the CDISC standards SDTM and ADaM mean that once these formats are adopted and used widely within companies there will be a unified market for reporting software for the first time in the pharma industry. Until now each company has had their own developed (more, or less) systems many with their roots in SAS V5 and relying on data _null_ for output. Their strengths are of course that they work and save programming. Their weaknesses tend to be documentation, brittleness (with concomitant poor error messages), restricted analysis datasets, and an inability to fully use metadata. Poor use of metadata means the same information may be entered in several places and therefore adds to the possibility of cross-output errors.

A major advantage of the CDISC formats is there is more metadata included – and standardised. This metadata includes so-called variable-level metadata which can be used to automate transpose operations and also to make them reversible without data loss.
This trend has already begun with the release of iCG from Insightful which uses ADaM datasets and the MAKS macros from KKZ Mainz which can report directly off SDTM and are available free. See the section 'Software' below for references.

SAS 9 IS NOT SAS 5

SAS has been significantly improved as a programming tool with the release of SAS 9. There are many useful functions and the availability of hash arrays and regular expressions take the data step to a new level. And the advent of JMP 7 with its SAS and stored process integration makes a new and powerful visual front end available.

armin 8/10/08 09:05

Comment [7]: Other front ends not less useful are SAS STAT STUDIO with V9 and Enterprise Guide (which not only can call Stored Processes, but is the preferred tool to create these).

GENETIC DATA AND OTHER MARKERS

These data are on the way and will need to be incorporated into patient subset definition or directly into tables and listings. This could be a huge amount of data (especially in Phase II while markers are still being assessed for utility) and just adding it to listings will be neither efficient nor feasible.

WILL IT REALLY CHANGE?

People have said statistical reporting must improve and change for at least 20 years, but I believe the pressures and opportunities are now coming together and there is a real opportunity now.

GETTING THE EMPEROR BACK IN HIS CLOTHES

READING AN EXAMPLE TABLE.

Here is a typical summary table. We are looking for differences from control. And we have confidence intervals, and it looks like the variability is uniform with time. We have to take the symmetry of the distribution on trust for now. However the eye can compare better when scanning up and down so the table arrangement is good for looking at time comparisons, but less good for comparing Active Drug and control.

	Active Drug				Control Drug			
Visit	N	Mean	SD	95% CI	N	Mean	SD	95% CI
1	112	118.1	1.3	(115.9, 120.3)	113	119.1	1.2	(117.0, 121.2)
2	112	122.7	1.4	(120.4, 125.0)	112	119.1	1.1	(117.0, 121.2)
3	111	125.6	1.0	(123.6, 127.6)	110	114.4	1.2	(112.3, 116.5)
4	110	133.1	1.2	(131.0, 135.2)	108	124.2	1.4	(121.9, 126.5)
5	110	136.7	1.2	(134.6, 138.8)	108	123.8	1.2	(121.7, 125.9)
6	108	134.2	1.1	(132.1, 136.3)	108	114.9	1.1	(112.8, 117.0)
7	106	131.0	1.2	(128.9, 133.1)	104	120.1	1.2	(118.0, 122.2)
8	105	126.2	1.3	(124.0, 128.4)	103	121.6	1.2	(119.5, 123.7)
9	104	124.3	1.2	(122.2, 126.4)	100	118.6	1.1	(116.5, 120.7)
10	102	125.1	1.2	(123.0,127.2)	100	117.7	1.3	(115.5, 119.9)

Figure 5 An example table of values for active drug vs control over ten visits (Soukup,2007)[28]

armin 8/10/08 09:08

Comment [8]: The screenshots of tables are referred to as "Figure" (like here), otherwise as "Table" as in Table 1. I find this a bit disturbing: Isn't it still a table (although it has been computer rasterized?

Here we are looking for differences in response over time. Not so easy to find even with confidence Intervals (CI). Now have a look at the graph In Figure 6.

This example is from Matt Soukup's excellent talk [28] and is one of the most dramatic examples I know showing how much more useful than tables are graphs for communication.

Tables are good for looking things up. Graphs make a difference to what we understand. This is not a small point, it is a big one. It is also true even for professionals trained in using tables as part of their daily work (see Gelman, Pasarica & Doghia [26]).

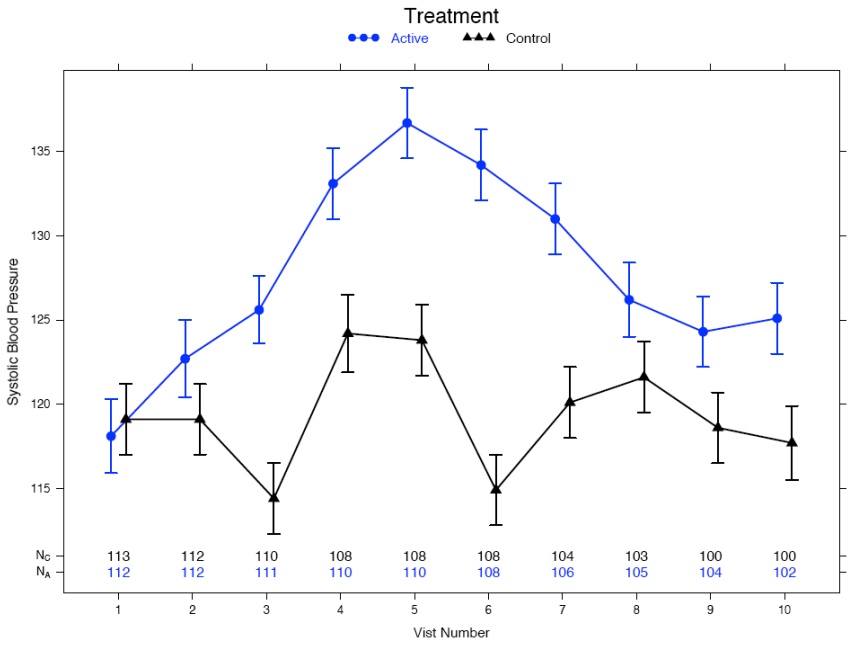


Figure 6 Plot of the data from Figure 5. The plot shows the treatment difference standing out dramatically. Soukup [28]

Let us look at some more examples of what is possible now.

ADVERSE EVENT DATA DISPLAY

Figure 7 shows us another table; this one is the top ten AEs from a Novartis study of Valsartan published on the clinical trials website and publicly available at
http://www.novartisclinicaltrials.com/webapp/clinicaltrialrepository/displayFile.do?trialResult=1928

Preferred term	Val/HCTZ 320/25 mg N=167 n (%)	Val/HCTZ 320/12.5 mg N=168 n (%)	Val/HCTZ 160/12.5 mg N=168 n (%)	Val 320 mg N=169 n (%)	Val 160 mg N=166 n (%)	HCTZ 25 mg N=165 n (%)	HCTZ 12.5 mg N=169 n (%)	Placebo N=169 n (%)	Total N = 1342 n (%)
Any adverse event	97 (58.1)	103 (61.3)	99 (58.9)	105 (62.1)	94 (56.6)	86 (51.8)	91 (53.8)	89 (52.7)	764 (56.9)
Dizziness	16 (9.6)	12 (7.1)	10 (6.0)	8 (4.7)	6 (3.6)	9 (5.4)	4 (2.4)	4 (2.4)	69 (5.1)
Cough	10 (6.0)	1 (0.6)	3 (1.8)	6 (3.6)	3 (1.8)	2 (1.2)	4 (2.4)	2 (1.2)	31 (2.3)
Upper respiratory tract infection	9 (5.4)	9 (5.4)	7 (4.2)	11 (6.5)	7 (4.2)	9 (5.4)	7 (4.1)	7 (4.1)	66 (4.9)
Arthralgia	8 (4.8)	1 (0.6)	3 (1.8)	5 (3.0)	4 (2.4)	0 (0.0)	4 (2.4)	3 (1.8)	28 (2.1)
Fatigue	7 (4.2)	4 (2.4)	5 (3.0)	6 (3.6)	2 (1.2)	6 (3.6)	5 (3.0)	3 (1.8)	38 (2.8)
Headache	7 (4.2)	10 (6.0)	11 (6.5)	12 (7.1)	12 (7.2)	13 (7.8)	14 (8.3)	22 (13.0)	101 (7.5)
Nasopharyngitis	7 (4.2)	15 (8.9)	11 (6.5)	5 (3.0)	14 (8.4)	4 (2.4)	8 (4.7)	4 (2.4)	68 (5.1)
Sinusitis	6 (3.6)	3 (1.8)	5 (3.0)	8 (4.7)	2 (1.2)	4 (2.4)	2 (1.2)	4 (2.4)	34 (2.5)
Nausea	5 (3.0)	3 (1.8)	1 (0.6)	7 (4.1)	2 (1.2)	6 (3.6)	0 (0.0)	3 (1.8)	27 (2.0)
Diarrhoea	4 (2.4)	5 (3.0)	5 (3.0)	4 (2.4)	4 (2.4)	2 (1.2)	4 (2.4)	3 (1.8)	31 (2.3)

Figure 7 Ten Most Frequent Reported Valsartan AEs overall by preferred term

This trial was unusual in that it had seven active drug treatments and Placebo. The treatments were combinations of Valsartan and HCTZ in various dose levels. The treatments form part of a 3x4 factorial design. In these data we are looking for trends across dose and differences from placebo and naturally it would be good to show that Valsartan had fewer AEs. This study is well controlled and so the patient numbers for each treatment are almost identical. This means we do not lose much by just looking at percentages. Nevertheless it is not easy to spot any trends here. The standard error of these differences depends on the total number and how close the percentage is to 50%. Not the easiest calculation to do in your head.

I am not suggesting we make formal tests here (for lots of reasons) but I am saying that a confidence interval is a much better calibrator of a difference than a difference of two percentages.

There is a technique for plotting AE incidences called the AE dotplot [31] and originated by Heiberger and Holland [14] and developed in the recent paper (Amit, Heiberger & Lane (2007)[6]).

First we can enter the data in a table like Figure 8 (with the fixed variable names) into a CSV file called aedotplot.dat with columns as described in Table 1.

Column	Variable name	**Content**
A	**RAND**	the treatment
B	**PREF**	the AE preferred term
C	**SN**	number of patients in that treatment
D	**SAE**	Number of patients with an AE of that preferred term.

Table 1 Data structure needed for ae.dotplot function

	A	B	C	D
1	RAND	PREF	SN	SAE
2	V-H 320/25mg	Dizziness	167	16
3	V-H 320/12.5mg	Dizziness	168	12
4	V-H 160/12.5mg	Dizziness	168	10
5	Val 320mg	Dizziness	169	8
6	Val 160mg	Dizziness	166	6
7	HCTZ 25mg	Dizziness	166	9
8	HCTZ 12.5mg	Dizziness	169	4
9	Placebo	Dizziness	169	4
10	V-H 320/25mg	Cough	167	10
11	V-H 320/12.5mg	Cough	168	1
12	V-H 160/12.5mg	Cough	168	3
13	Val 320mg	Cough	169	6
14	Val 160mg	Cough	166	3
15	HCTZ 25mg	Cough	166	2
16	HCTZ 12.5mg	Cough	169	4
17	Placebo	Cough	169	2
18	V-H 320/25mg	Upper respiratory t	167	9

Figure 8 Sample data table entered from Figure 7 ready to be used by the ae.dotplot function

An R program to make an aedotplot from data with just one treatment and placebo (taken from the HH package documentation [31]):

```
# Read the data from a file in the current directory
aeanonym <- read.table("aedotplot.dat", header=TRUE, sep=",")
     aeanonym[1:4,]     # the data we need are in the first 4 columns
## Calculate log relative risk and confidence intervals (95%)
## logrelrisk sets the sort order for PREF to match the relative risk.
aeanonymr <- logrelrisk(aeanonym,
   A.name=levels(aeanonym$RAND)[1],
   B.name=levels(aeanonym$RAND)[2])
aeanonymr[1:4,]
## construct and print plot on current graphics device
ae.dotplot(aeanonymr,
   A.name="Placebo",
   B.name="Val 320mg")
```

The program reads the data from a CSV file calculates the log relative risk (logrelrisk) (and sorts by it) and then makes the two-panel plot. The result is plotted by calling print on the aedotplot object. For comparing treatment Valsartan (Diovan®) 320mg vs Placebo we have Figure 9.

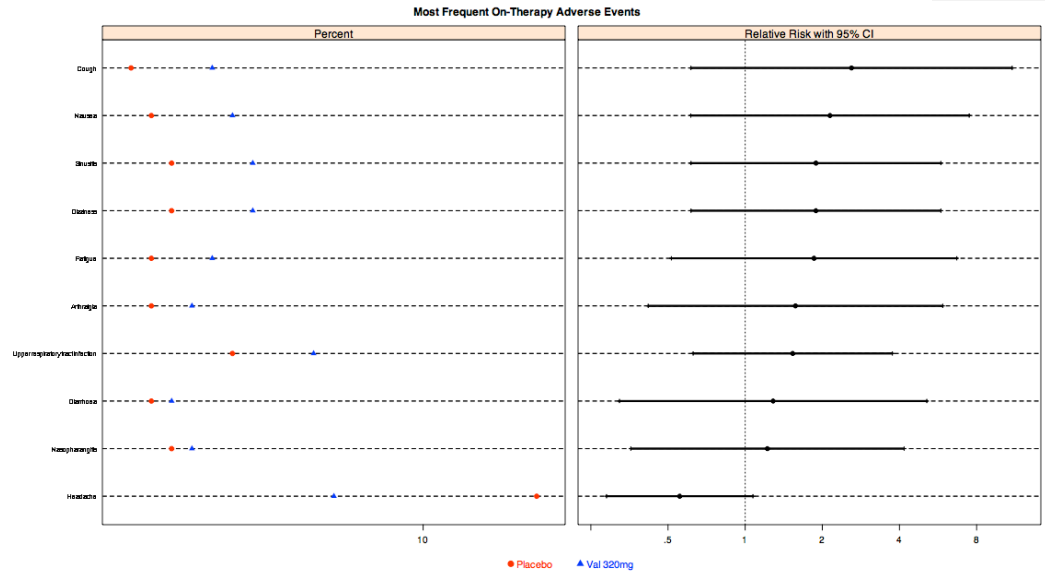


Figure 9 AEdotplot for data from Valsartan trial CVAH631C2301, percentage of patients reporting the AE in left panel and relative risk of AE in high dose group vs. Placebo in the right panel.

The AEdotplot [31] function uses a two panel display – on the left is a dot plot of the percentages calculated from and on the right hand panel is the relative risk and its 95% CI (plotted on a log scale). The plot is sorted by relative risk, with the highest at the top. The relative risk is related to the gap between the percentages on the left panel. There is a clear pattern visible now, first from the right-hand panel we can see there are no AEs with strong evidence they are more common in the high dose group vs Placebo. This conclusion is not really accessible from the table. Second, the data for 'Headache' show a different pattern from the other AEs, it has been included in the top ten because it is a common AE, but it is actually *less* common in the high dose group than in placebo. The difference is close to 'significance'.

This finding suggests we look at other treatments as well and the results are in Figure 10. The dot plot makes it easy to notice the pattern is different for 'Headache' than the other AEs. For this AE the order of the (red) dots and (blue) triangles is reversed and the difference is largest for the combination treatment. In contrast 'Dizzyness' shows a pattern of increase with dose.

This is not a paper about Valsartan adverse events so I will not go any further with comparisons here but I will note that the labeling for Valsartan available at http://www.inhousedrugstore.co.uk/heart-health/valzaar.html states:

> **Side Effects**
>
> Valsartan may cause side effects. Tell your doctor if any of these symptoms are severe or do not go away:
>
> Dizziness, ***headache***, excessive tiredness, diarrhea, stomach pain, back pain, joint pain

These are of course only the data from one trial and so we should not jump quickly to conclusions, nevertheless, the value of an analytical-graphical analysis is clear.

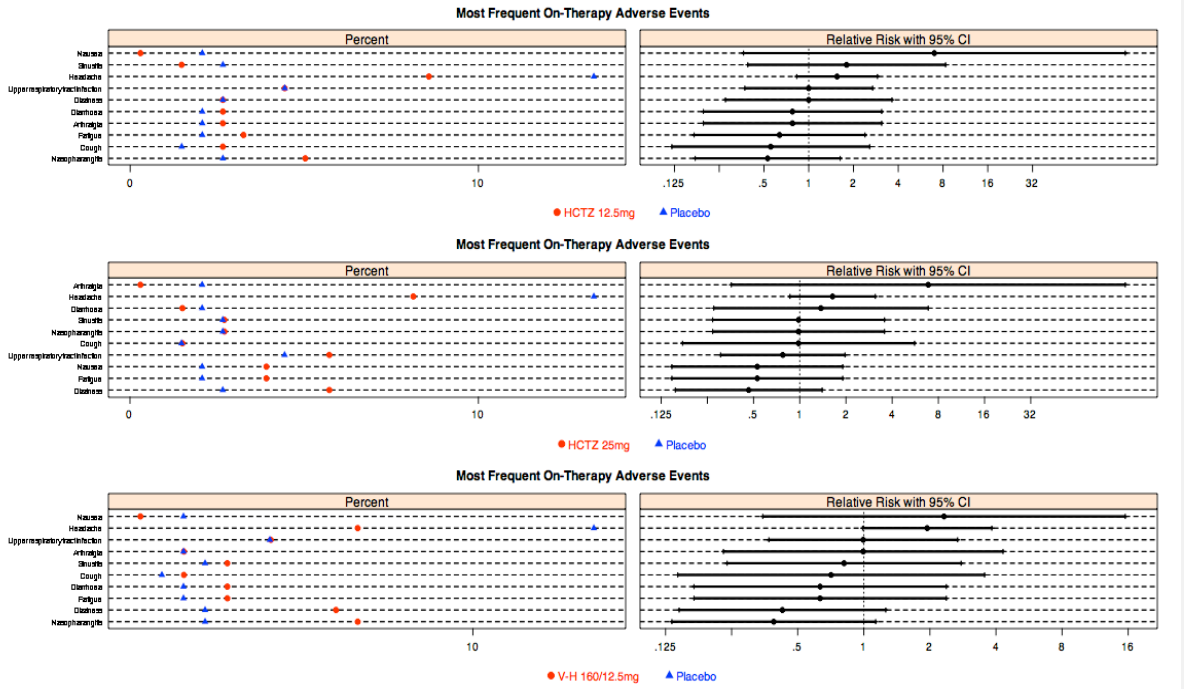


Figure 10 Series of AE dotplots designed for comparing multiple treatments. Three treatments are shown The two HCTZ treatments and one Valsartan-HCTZ combination.

The plots above were done using R and the HH package [28] [13] but they are also possible with other tools, such as JMP®.

A sample AE dotplot made with JMP, using different data, is shown in Figure 11. This is the code [Meintraub, *Pers. Comm.*]:

```
Clear Globals();
Clear Log();
::dt = Current Data Table();
::Max_per = Col Max( Column( "Max value" ) );
::Max_RR = Round( Col Max( Column( "RR CI up" ) ), -1 ) + 10;
::cc1 = Chart(
    X( :Adverse Reaction ),
    Y( :perc A, :perc B, :Max Value ),
    Horizontal( 1 ),
    Overlay Axis << {{Scale( Linear ), Format( "Fixed Dec", 0 ),
    Min( 0 ), Max( ::Max_per )}},
    Y[1] << Point Chart( 1 ),
    Y[2] << Point Chart( 1 ),
    Y[3] << {Needle Chart( 1 ), Show Points( 0 ), Overlay Color( 32 )
    }
);
::rcc1 = ::cc1 << report;
::pb1 = ::rcc1[Picture Box( 1 )];
```

```
::rcc1[Text Edit Box( 1 )] << Set Text( "Percent" );
::cc2 = Chart(
    X( :Adverse Reaction ),
    Y( :Relative Risk, :RR CI low, :RR CI up ),
    Horizontal( 1 ),
    Category Axis << {Label( None ), Axis Name( " " )},
    Overlay Axis << {{Scale( Log ), Format( "Best" ), Min( 0.1 ),
    Max( ::Max_RR ), Inc( 1 ), Minor Ticks( 8 )}},
    Range Chart( 1 ),
    Y[1] << {Show Points( 1 ), Overlay Marker( 12 )},
    SendToReport(
        Dispatch(
            {},
            "107",
            ScaleBox,
            {Scale( Log ), Format( "Best" ), Min( 0.1 ),
            Max( ::Max_RR ), Inc( 1 ), Minor Ticks( 8 ),
            Add Ref Line( 1, Dotted, Black )}
        )
    )
);
```

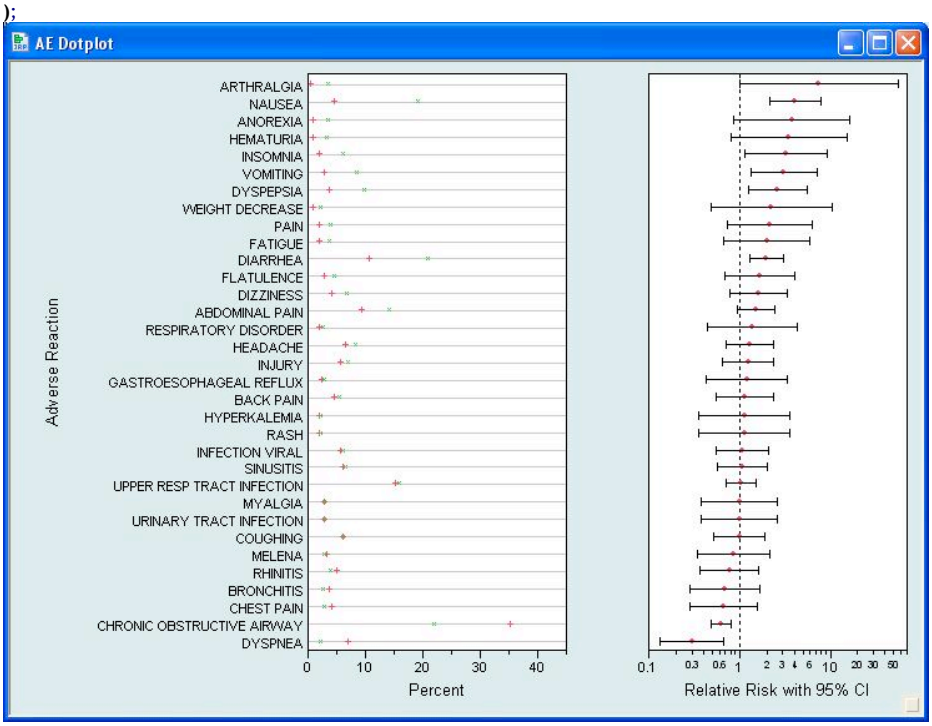


Figure 11 Two panel AE dotplot created with JMP

```
::rcc2 = ::cc2 << report;
::pb2 = ::rcc2[Picture Box( 1 )];
```

```
::rcc2[Text Edit Box( 1 )] << Set Text( "Relative Risk with 95% CI" );
New Window( "AE Dotplot", H List Box( ::pb1, ::pb2 ) );
::rcc1 << Close Window();
::rcc2 << Close Window();
```

Note that the JMP script has not been packaged as a function like AEdotplot so it should not be compared directly to the R code above.

Examining particular AEs

The above analyses though useful have actually discarded a lot of data. We have only examined the incidence of an adverse event per patient. We have discarded all the information about recurrence, severity and time of occurrence. When we need to examine particular AEs we can use the powerful statistics developed for time to event data and not discard so much information. In Figure 12, we compare the time since randomization to experience the event for two treatments. This plot is readily available in the new Insightful Clinical Graphics package although this figure is taken from [6].

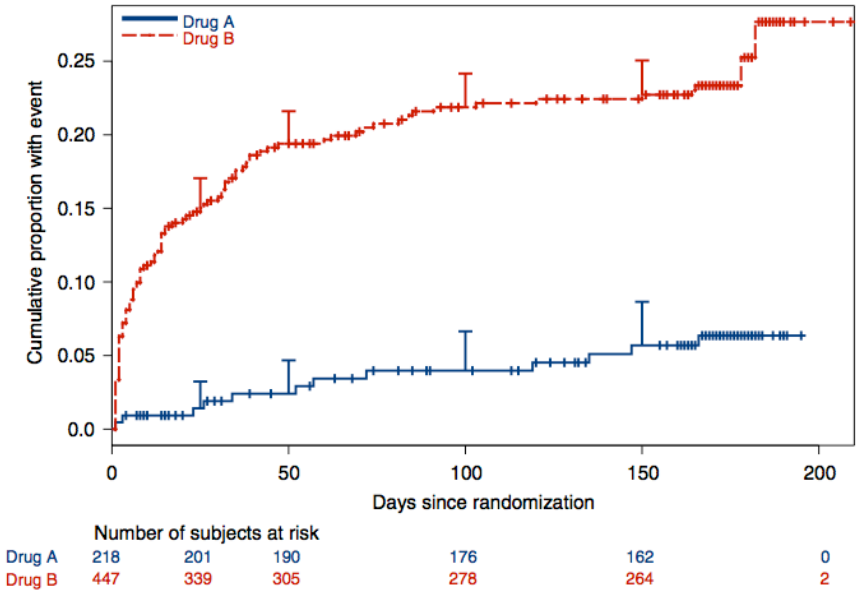


Figure 12 Cumulative distribution (with SEs) of time to first AE of special interest.[6]

Here there is a much higher risk of the AE for drug B, this can also be shown by plotting the hazard function which is shown in Figure 12. A figure of cumulative proportion tells a whole story but it is not so clear at what time points the risk is changing most. This can be seen clearly from the hazard function (estimate) plot in Figure 13 where it is clear that the differences lie in the first 40 days of treatment. After that period the relative risks of the AE for drug A and B are not distinguishable.

Although I have not included the AE table here it is clear how the graphics really expose the issues of interest. This kind of analysis is just not possible from a table of incidences.

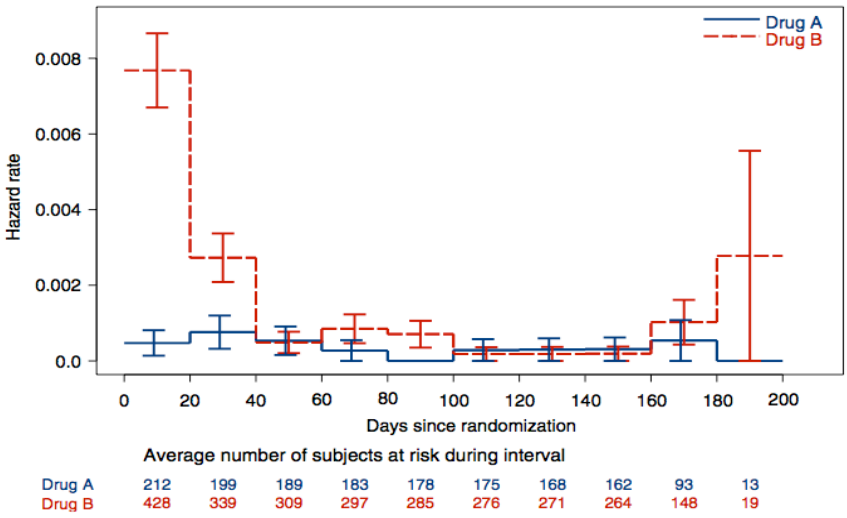


Figure 13 Hazard function for an AE of special interest (with SEs)

LABORATORY DATA

The data for liver enzymes are very variable but also very important to assess. For this job we need box plots because of the variability, asymmetry and importance of the few high values. Again a high resolution plot gives much better valueFigure 14 illustrates this with a plot from Heiberger and

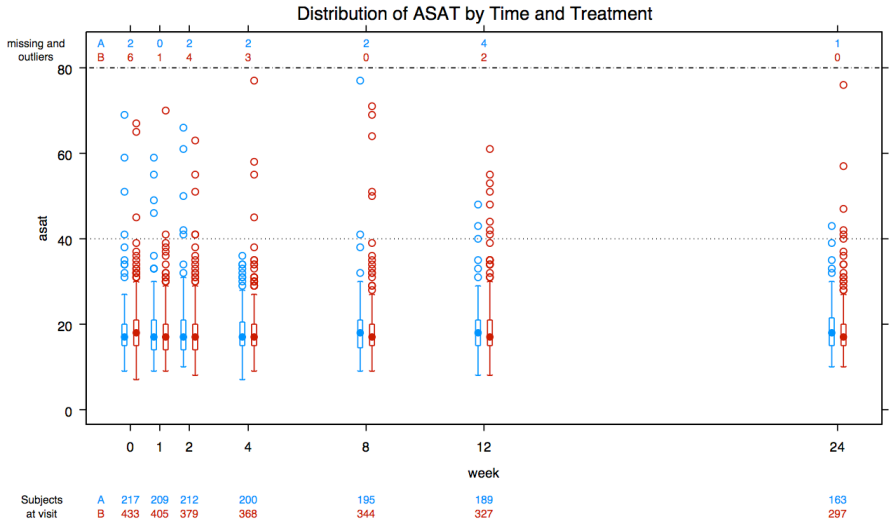


Figure 14 Coloured Boxplot showing distribution of ASAT by time and treatment

Holland (2004)[13, 29]. Here it is important to scale the X axis by time and not by conventional

armin 8/10/08 09:13

Comment [9]: I would only refer to as Figure 14, not with full text description.

armin 8/10/08 09:15

Comment [10]: There are some "impurities" at the top of the graphic, due to the screenshot not precisely placed.

visit number, and to show the number of missing values. The range of the graph is also restricted because there are a very few exceptionally high values and including them would compress the Y axis and make detail in the lower range invisible. The numbers of excluded outliers and missing value is given for each time point along the top of the graph. But we are looking just one parameter and we discussed above that this is not enough.

MULTIVARIATE DISPLAYS OF LIVER ENZYME DATA

armin 8/10/08 09:16

Comment [11]: In this section it sometimes says ALT and AST, sometimes ALAT and ASAT. The terminology should be uniform.

The analysis of Liver function measurements (LFTs) is an inherently multivariate one and displays are available that take this into account. The essential questions are:

- Do ALT (ALAT) and AST (ASAT) track together?
- Are there simultaneous elevations in ALT/AST and Bilirubin?
- What is the time-course of the elevations?

These questions derive from the well known Hy's law which gives rules of thumb relating LFT results to liver damage. This shift plot illustrates this very well in a useful plot combining technique of lattice graphics with the shift table as discussed in Amit et al. [6].

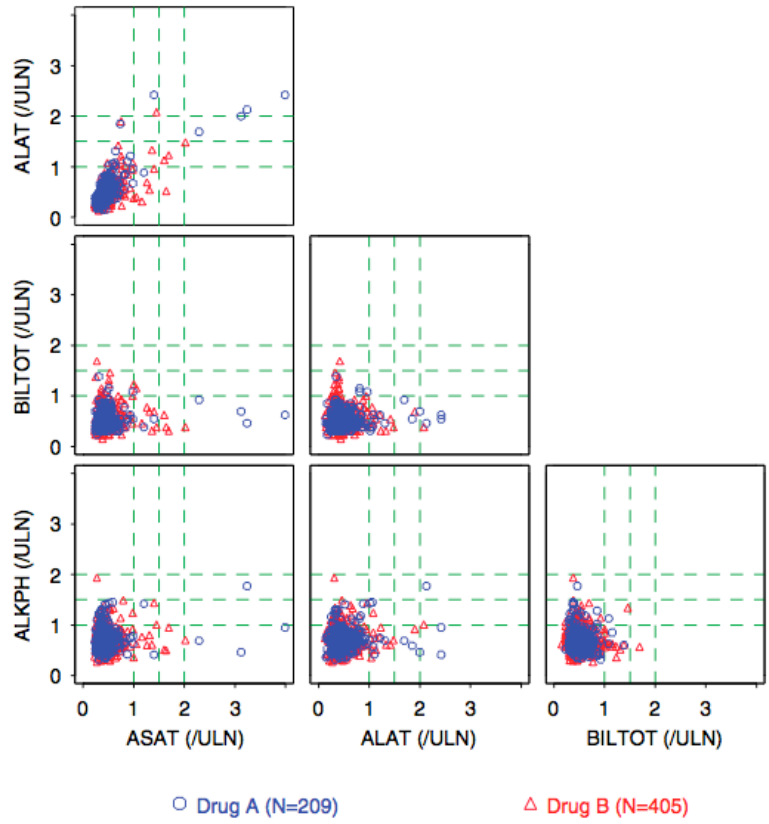
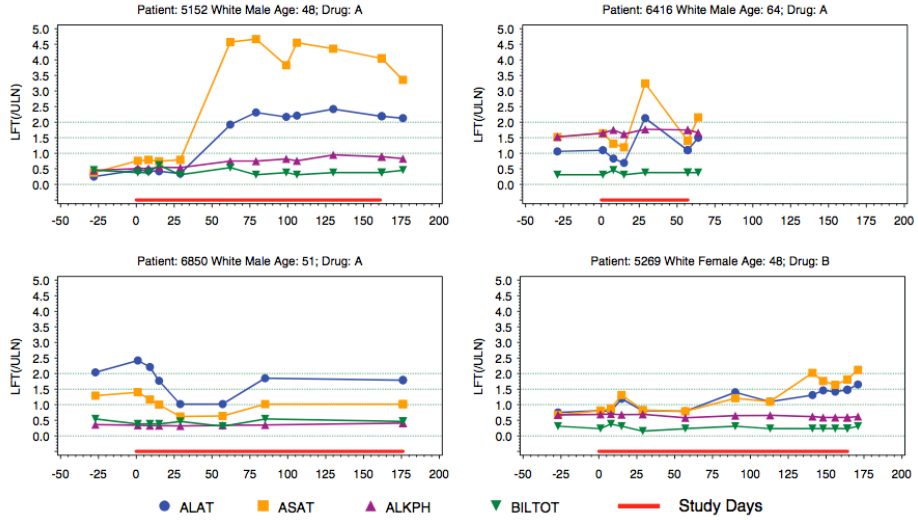


Figure 15 Matrix display of shift from baseline to maximum LFT values per patient. [6]

In Figure 15 there are four outlying values of ALT/ASAT (not all above the limits of concern in

one dimension) and the above plot is really a tool to find which patients need to be looked at in detail. Their plots are shown Figure 16 and we see very different patterns of response over time. Patient 6850 (bottom left quadrant) actually improves after drug starts. A fuller investigation of these cases can now be done and would include checking for concomitant medication with known hepattoxic drugs, and checking what reason patient 6416 (top right quadrant) withdrew from the study.



For ALAT, ASAT and ALKPH, the Clinical Concern Level is 2 ULN;
For BILTOT, the CCL is 1.5 ULN; where ULN is the Upper Level of Normal Range

Figure 16 Time series plots of LFT data from four patients

At this point we would like to be able to state these are the only four patients that could have these problems but we cannot be sure about that because we have ignored data and also because of patient 5269 in Figure 16 (bottom right quadrant) shows a gradual onset of increased ASAT/ALAT. There could be other patients with a similar pattern which do not happen to reach quite the extreme values that patient 5269 does. We have not searched for this pattern within the 'normal' patients. Techniques for doing this search still need to be refined and this is an interesting area for further work.

The new iCG package from Insightful has a variety of this plot and can mark individual points as being ones violating Hy's law.

It also has a novel model for classifying changes in lab values as treatment emergent. This uses the arbour/ forest library in S-Plus and looks like a very powerful way to diagnoses general issues with lab parameters. The model is introduced here: http://en.wikipedia.org/wiki/Random_forest An R package is documented in the R newsletter here: http://cran.r-project.org/doc/Rnews/Rnews_2002-3.pdf

SUMMARY:

Safety reporting is becoming more important to drug development and big improvements already exist and can be done with modern tools. There are two directions for improvements – first using more graphics to communicate the data, and second more analytical approaches that put the sound bases behind those plots. The perfect methods of analysis and display for each kind of safety data have not been found yet, so there is a lot of interesting work to be done.

REFERENCES

[1] Shi-Tao Yeh, A SAS Macro For Producing Clinical Laboratory Shift Table, http://www.lexjansen.com/pharmasug/2003/posters/p111.pdf

[2] Trost, DC. Multivariate probability-based detection of drug-induced hepatic signals. Toxicol Rev. 2006;25(1):37-54

[3] Merz, M. Spotting clinical safety signals earlier: the power of graphical display. pdf at : http://spotfire.tibco.com/spotfire_downloads/customer_presentations/uc2006/michael_merz.pdf

[4] Mold JW, Aspy CB, Blick KE, Lawler FH. The determination and interpretation of reference intervals for multichannel serum chemistry tests. J FAM PRACT. 1998 Mar;46(3):233-41.

[5] Richard C. Friedberg, MD, PhD; Rhona Souers, MS; Elizabeth A. Wagar, MD; Ana K. Stankovic, MD, PhD, MPH; Paul N. Valenstein, MD. The Origin of Reference Intervals,. *Archives of Pathology and Laboratory Medicine:* Vol. 131, No. 3, pp. 348–357. PDF from: http://arpa.allenpress.com/pdfserv/10.1043%2F1543-2165(2007)131%5B348:TOORI%5D2.0.CO%3B2

[6] Ohad Amit, Richard M. Heiberger, and Peter W. Lane, (2007),'Graphical Approaches to the Analysis of Safety Data from Clinical Trials', Pharmaceutical Statistics, Published Online: 26 Feb 2007, http://www3.interscience.wiley.com/cgi-bin/abstract/114129388/ABSTRACT

[7] John R Tukey, Exploratory Data Analysis, Addison Wesley, (1977) Publ. Co., 1985

[8] William S. Cleveland, Visualizing Data, Hobart Press, 1993 and William S. Cleveland, The Elements of Graphing Data, Wadsworth

[9] E. R. Tufte, The Visual Display of Quantitative Information, Cheshire, CT: Graphics Press, 1989 (and three other books, see http://www.edwardtufte.com/tufte/ for more).

[10] Frank Harrell's Course on statistical graphics is here : http://biostat.mc.vanderbilt.edu/twiki/pub/Main/StatGraphCourse/graphscourse.pdf

[11] Wilkinson, Leland, Presentation Graphics, in International Encyclopaedia of the Social & Behavioural Sciences, (2001) 26 vols. Oxford: Elsevier

[12] Rreport: Source code and documentation obtainable with sample reports from the website http://biostat.mc.vanderbilt.edu/twiki/bin/view/Main/Rreport

[13] Richard M. Heiberger (2008). HH: Statistical Analysis and Data Display: Heiberger and Holland. R package version 2.1-12.

[14] Heiberger, R. and Holland, B. (2004) Statistical Analysis and Data Display: an Intermediate Course with Examples in S-PLUS, R, and SAS. Springer-Verlag, NY. http://springeronline.com/0-387-40270-5

[15] M. Friendly, Graphical methods for Categorical Data, SAS User Group International Conference Proceedings, 17:190-200, 1992.

[16] Tukey, J. W. and Tukey, P. A. (1985). Computer graphics and exploratory data analysis: An introduction. In Proceedings of the Sixth Annual Conference and Exposition: Computer Graphics'85 3 773-785. National Computer Graphics Association, Fairfax, VA.,

[17] Wilkinson, L., Anand, A., Grossman, R., Graph-theoretic scagnostics, Information Visualization, 2005. INFOVIS 2005, IEEE Symposium on Volume, Issue , 23-25 Oct. 2005 Page(s): 157 – 164.
Pdf available at http://www.rgrossman.com/dl/proc-094.pdf

[18] Wilkinson L, Anand, A and Grossman, R., High-dimensional Visual Analytics: Interactive Exploration Guided by Pairwise Views of Point Distribution, IEEE Transactions on Visualization and Computer Graphics, Volume 12, Number 6, pages 1363-1372, 2006. http://www.rgrossman.com/dl/journal-033.pdf .

[19] Heike Hofmann, Lee Wilkinson, Hadley Wickham, Duncan Temple Lang and Anushka Anand; The Scagnostics R package. available on CRAN at http://cran.r-project.org/web/packages/scagnostics/index.html

[20] Carr, D. & Lewis Koh, N., N, The Hexbin R package at http://bioconductor.org/packages/2.2/bioc/html/hexbin.html
[21] Hyndman, R. and Einbeck, J., The hdrcde package http://cran.r-project.org/web/packages/hdrcde/index.html
[22] Venables, W.N. and Ripley, B (1999). Modern Applied Statistics with S, 4th ed. (2002), Chapter 9 - Tree-Based Methods.
[23] Berry SM, Berry DA (2004), Accounting for multiplicities in assessing drug safety: A three-level hierarchical mixture model, *Biometrics* 2004; 60(2):418-426.2004
[24] LI, Q., Bayesian Inference On Dynamics Of Individual And Population Hepatotoxicity Via State Space Models, PHD Thesis, 2005, Ohio State University.
[25] Asur, Detecting Hepatotoxicity in Clinical Trials (2006)
[26] A. Gelman, Pasarica C & Doghia R - Practice what we preach?, Tables into graphs. Statistical computing and Graphics (2002) http://www.stat.columbia.edu/~gelman/research/published/dodhia.pdf
[27] Kastellac & Leone, Tables to graphs- http://www.columbia.edu/~jpk2004/graphs.pdf and website: http://tables2graphs.com/doku.php
[28] Soukup, M., Visual Representation of Clinical Data to Elicit Safety and Efficacy Signals, DIA Congress 2007. Pdf available from http://www.insightful.com/insightful_doclib/document.asp?id=417
[29] Wang, J., Using graphics to discover and explore. http://stat-computing.org/events/2007-jsm/wang.pdf
[30] O'Connell, M, Standardized Graphics for Safety using S-PLUS Software (2006) http://bass.georgiasouthern.edu/PDFs/BASS%202006%20OConnell.pdf
[31] Aedotplot documentation at: http://bm2.genes.nig.ac.jp/RGM2/R_current/library/HH/man/ae.dotplot.html
[32] Figure 14 from http://astro.temple.edu/~rmh/HH/bwplot-color.pdf

ACKNOWLEDGMENTS

Many thanks to Richard Heiberger for help with Figure 9 and Figure 10 and to David Meintraub for help with Figure 11. Thanks also to those others who gave me permission to reproduce their figures. I would also like to thank Business & Decision for funding some of my time working on this paper.

armin 8/10/08 09:17
Comment [12]: Not bold.

SOFTWARE

iCG : http://www.insightful.com/industry/pharm/clinicalgraphics.asp
Patient Profilers - :

- PPD http://www.csscomp.com/web/products/patientprofiles.htm
- Free (and html, not graphic): http://www.datasavantconsulting.com/roland/rgpp.html
- Phase Forward: http://www.phaseforward.com/products/safety/aer/

R: http://www.r-project.org/ For a summary of graphic options see http://cran.r-project.org/web/views/Graphics.html
JMP: http://www.jmp.com/
SAS Macros for SDTM datasets from KKZ-Mainz report see last years talk at PhUSE http://www.lexjansen.com/phuse/2007/ad/ad04.pdf Email Daniel Wachtlin wachtlin[at]izks-mainz.de for a copy and note the address in the PhUSE paper.
Spotfire life science gallery.
http://spotfire.tibco.com/community/blogs/lifesciencesgallery/pages/dxp-clinical-labs-mini-solution.aspx

armin 8/10/08 09:19
Comment [13]: Remove "mailto:"

CONTACT INFORMATION

I would value your comments and questions on this paper. Please contact the author at:

David J Garbutt
Business & Decision
Löwenstrasse 12
CH – 8001 Zürich
Work Phone: +41 44 390 37 21
Fax: +41 44 390 3722
Email: david.garbutt@businessdecision.com
Web: http://www.businessdecision.ch/2302-life-sciences-consulting-services.htm