

Extreme value modelling of clinical trial safety data

Harry Southworth PhD
Data Clarity Consulting Ltd
harry@dataclarityconsulting.co.uk

Janet E. Heffernan PhD
J. Heffernan Consulting
jan@jheffernanconsulting.com

Overview

- ▶ Specification of the problem
- ▶ Introduction to extreme value modelling
- ▶ Univariate analysis
- ▶ Multivariate analysis

Safety data from clinical trials

- ▶ Second most common reason for late phase failure is **safety**
- ▶ Most common reason for market withdrawal is **safety**
- ▶ Why?
 - ▶ Trials designed to address a well-specified efficacy question
 - ▶ Most of the data relate to safety
 - ▶ Adverse events, lab data, vital signs, etc.
 - ▶ A mixture of binomial, time-to-event, and continuous but highly non-Gaussian data
 - ▶ Impossible to avoid interest in data driven questions
 - ▶ Leads to concern over **lack of pre-specification, multiplicity, power**

Expected values can't help

- ▶ Most statistical methods are designed to predict **expected values**
 - ▶ Mean, median, regression models
- ▶ With safety, it's usually the **unexpected values** that are of interest
 - ▶ Unusually big or small values in labs, vital signs
 - ▶ There is no reason to suppose that the behaviour of the data near the centre of the distribution is informative about values in the tails
 - ▶ To characterize the centre and assume it tells us something about the safety of the drug can be misleading and dangerous
- ▶ So, **the usual statistical methods can't help!**

Extreme values of ALT

- ▶ The rest of the presentation focusses on extremes of **liver related lab variables**
 - ▶ Large values of ALT suggest potential liver injury
 - ▶ Potential for liver injury has been the most frequent reason for safety related withdrawal of drugs from the market [1]
 - ▶ *ticrynafen, benoxaprofen, bromfenac, troglitazone, nefazodone, ximelagatran...*

Extreme values of ALT (2)

- ▶ According to published guidance (CTC [3]):

| | Grade | Severity |
|--|-------|-------------------------|
| $ULN \leq ALT < 2.5 \times ULN$ | 1 | Mild |
| $2.5 \times ULN \leq ALT < 5 \times ULN$ | 2 | Moderate |
| $5 \times ULN \leq ALT < 20 \times ULN$ | 3 | Severe |
| $ALT > 20 \times ULN$ | 4 | Life threatening |

- ▶ ULN can be thought of as the units of measurement for ALT

Example: troglitazone

Troglitazone (for treatment of diabetes). FDA review states

Mean [ALT] levels fell in patients receiving troglitazone in phase 3 trials... *It was also stated that 2.2% of patients in phase 3 trials had an [ALT] level exceeding $3 \times ULN$... What was not appreciated by [FDA] was that many of the patients classified as $ALT > 3 \times ULN$ actually had ALT values that **were VERY much higher** than $3 \times ULN$... 23 patients had treatment-emergent ALT values over $3 \times ULN$... In 14 of these 23 patients, the ALT value exceeded $8 \times ULN$... and in 5/23 patients the **ALT value exceeded $30 \times ULN$.***

The drug was withdrawn from the market after reports of liver failure and death

ALT isn't the full story

Hy's Law

So far as I (Harry) understand what I've read...

- ▶ If ALT or AST is 'high', then
- ▶ If total bilirubin (TBL) is 'high', then
 - ▶ If alkaline phosphatase (ALP) is normal
 - ▶ **Conclude drug induced liver injury**
 - ▶ If ALP is 'high'
 - ▶ **Conclude underlying liver disease**

Which logically implies that if we only develop drugs that cause ALP elevation, we **cannot cause liver injury**

- ▶ ... which seems problematic

Also...

According to the Health Canada guideline [2]

- ▶ 'It has been noted that $ALP > 2 \times ULN$ **occurs in one third of potential Hy's Law cases** and can be associated with subsequent liver failure'

which goes uncited but presumably comes from Kaplowitz [4] (presentation at US FDA liver injury conference, 2010)

So the interpretation of data is

- ▶ Multivariate
- ▶ Disputed

Extreme value modelling

- ▶ A mature branch of statistics
 - ▶ Key publication by R. A. Fisher and L. H. C. Tippett, **1928**
 - ▶ First full length textbook by E. Gumbel, **1958**
- ▶ Used in many areas of application:
 - ▶ Meteorology, insurance, finance, geology, metallurgy, materials science, network traffic, ...

We wish to describe the tails of a distribution

- ▶ **extrapolation** – at your own risk!!
- ▶ parametric model needed
- ▶ calibrate using **data from tails**

... but something **more reassuring** is needed.

Asymptotically motivated statistical models

Recall... **Central Limit Theorem**

- ▶ mean \bar{X}_n of IID observations X_1, \dots, X_n mean μ , finite variance σ
- ▶ $\bar{X}_n \rightarrow \mu$ in probability
- ▶ normalised \bar{X}_n

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \rightarrow \mathbf{Normal}(0, 1) \text{ as } n \rightarrow \infty,$$

Asymptotic argument for choice of statistical model

Ingredients:

- ▶ IID data (distribution unknown)
- ▶ operation induces stability (for CLT this is **SUMMING**)
- ▶ normalisation to avoid degeneracy
- ▶ limit distribution in parametric family (for CLT this is **NORMAL**)

Limit argument used for finite samples when n is "**sufficiently large**".

Our interest is in the tails of the distribution

Do the same as in the CLT but use **Maximum** instead of **Mean**

This gives us the **Extremal Types Theorem**

- ▶ $\max M_n$ of IID observations X_1, \dots, X_n
- ▶ $M_n \rightarrow$ **upper end point** of distribution in probability
- ▶ normalised M_n – if limit exists then

$$\frac{M_n - a_n}{b_n} \rightarrow \text{GEV}(\xi) \text{ as } n \rightarrow \infty,$$

GEV is the **Generalised Extreme Value Distribution**.

Model for threshold excesses

- ▶ use the **GEV for maxima**
- ▶ derive model for tails of distribution F of X_1, \dots, X_n

Generalised Pareto distribution

If assumptions behind GEV hold then:

- ▶ there exists **threshold** u excesses of which are

$$Y_u = (X - u)_+.$$

- ▶ then $Y_u | Y_u > 0 \sim \text{GPD}(\sigma_u, \xi)$

$$\Pr(Y_u < y | Y_u > 0) = 1 - [1 + \xi y / \sigma_u]_+^{-1/\xi},$$

The generalized Pareto distribution

Two parameters:

- ▶ σ_u : the scale parameter
- ▶ ξ : the shape parameter
 - ▶ $\xi < 0$: short tailed distribution (finite upper end point)
 - ▶ $\xi = 0$: **exponential** distribution
 - ▶ $\xi \geq 0$: heavy tailed distribution
- ▶ **no straightforward physical meaning**
 - ▶ Results often presented using
 - ▶ Predicted probabilities of exceeding certain thresholds
 - ▶ Predicted extreme quantiles of the distribution – **return levels**

Univariate results

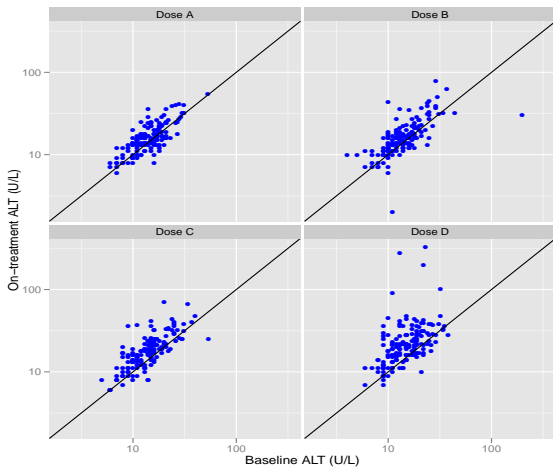
Data from a single clinical trial

- ▶ Patients randomized to one of four doses of drug
 - ▶ **Doses coded A (lowest), B, C, D (highest)**
- ▶ Approximately **160 patients per dose**
- ▶ A baseline assessment of ALT, AST, TBL, ALP
- ▶ And a single on-treatment assessment

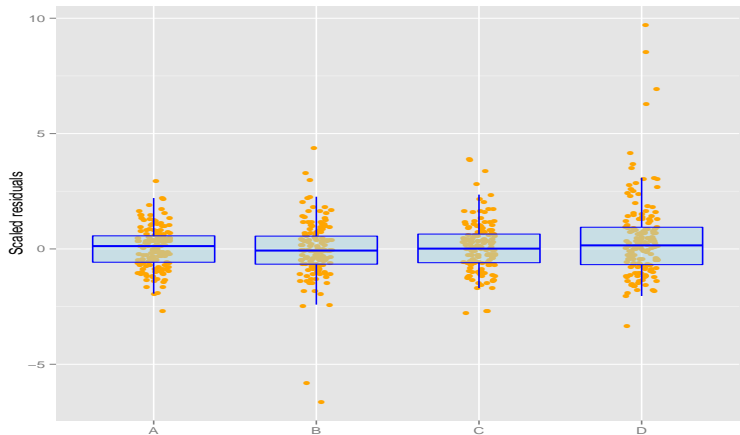
The highest dose of the drug has been associated with liver injury in the literature

First, let's look at ALT alone

Shiftplots of ALT



Boxplots of residuals after removing baseline effect



GPD models for ALT

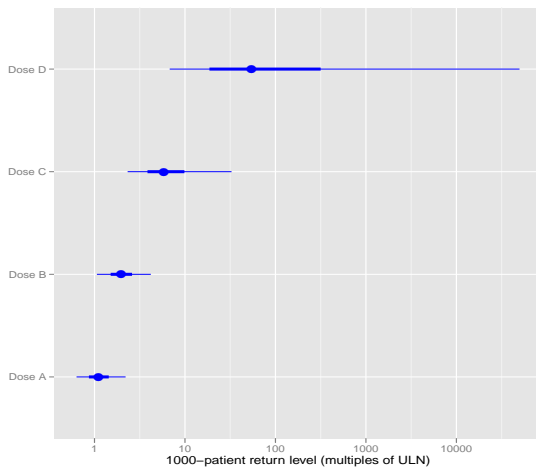
Consider models with term for dose in

- ▶ $\phi = \log \sigma$
- ▶ ξ
- ▶ ϕ and ξ

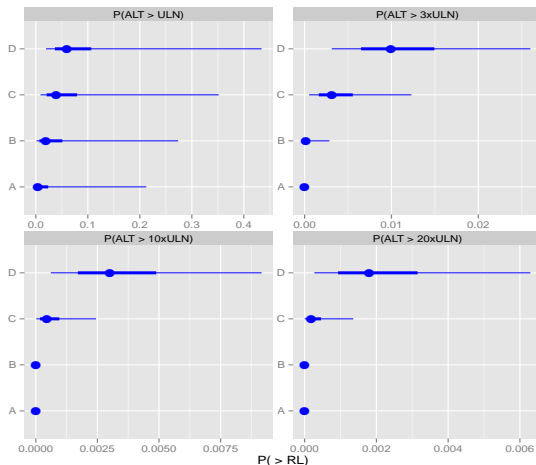
Choose between models using Akaike's Information Criterion

- ▶ Model with linear term for dose in ξ is preferred
- ▶ $\hat{\xi} < 0$ for doses A and B, $\hat{\xi} > 0$ **for C and D**

Predicted 1000-patient return levels



Predicted probabilities of threshold exceedance



Multivariate extremes

What is a multivariate extreme value?

- ▶ componentwise **maxima**?
- ▶ points that are **large** in
 - ▶ **all** components?
 - ▶ **any** component?
 - ▶ **a given** component?

A Conditional Model for Multivariate Extremes

Consider multivariate response (X, Y) .

- ▶ work with standardised margins
- ▶ look at

$$\lim_{u \rightarrow \infty} \Pr(Y < y | X > u)$$

limiting form of conditional distribution as threshold $u \rightarrow \infty$.

- ▶ if X and Y **are at all dependent** then for any fixed y this limit will be **degenerate**
- ▶ need to **center and scale** Y to track its growth with X

A Conditional Model for Multivariate Extremes

Standardised Laplace margins are convenient (the maths works!)

Then the appropriate centering and scaling is:

$$\frac{Y - ax}{x^b}$$

for large values x of conditioning variable X .

There are two model parameters:

- ▶ a describes change of **location** with threshold
- ▶ b describes change in **scale** with threshold

Regression interpretation

Defining

$$Z = \frac{Y - ax}{x^b}$$

we can think of this as a regression model:

$$Y = ax + x^b Z$$

Z are **residuals**, assumed independent of **explanatory variable** x .

Detailed model specification

Marginal model for each variable

- ▶ **GPD tail** above threshold
- ▶ **empirical** distribution below threshold
- ▶ used to **transform margins** to standard (Laplace) margins

Dependence model

- ▶ threshold model – **fit to data for which $X > v$** for high v
- ▶ **parameters a and b** capture dependence of Y on X
- ▶ residuals Z assumed **independent** of X above threshold

Sample based extrapolation

Sample from distribution of (X, Y) **given** $X > t$ for $t > v$

1. simulate value x from **tail** of $X | X > t$
2. sample from **observed model residuals** Z
3. construct

$$Y = ax + x^b Z$$

4. (X, Y) is an **importance sample** from the distribution of

$$(X, Y) | X > t$$

Multivariate results

From the clinical trial described previously, look at **TBL and ALP as well as ALT**

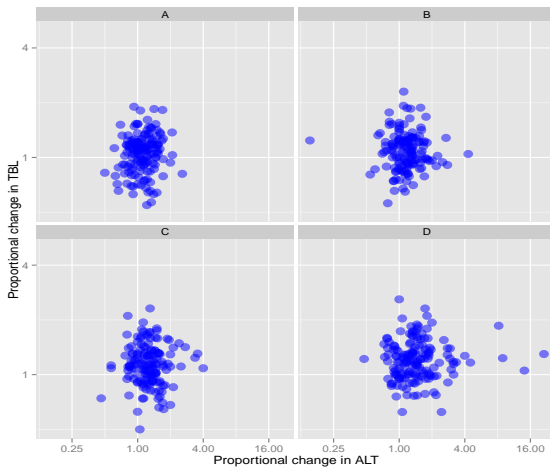
- ▶ Ignore AST to simplify output

Remember:

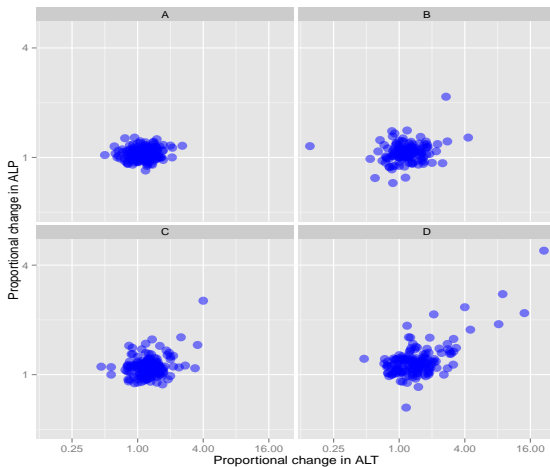
- ▶ The highest dose has been associated with liver injury in the literature
- ▶ Traditional wisdom has it that liver injury results in **high ALT and TBL, but not high ALP**

Here we'll focus on **proportional change from baseline**

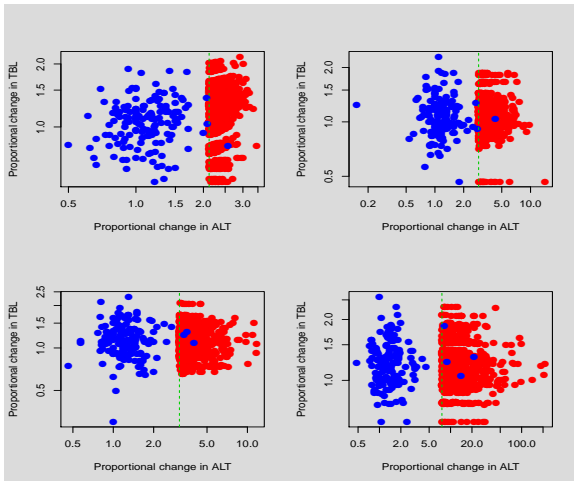
Observed TBL vs ALT



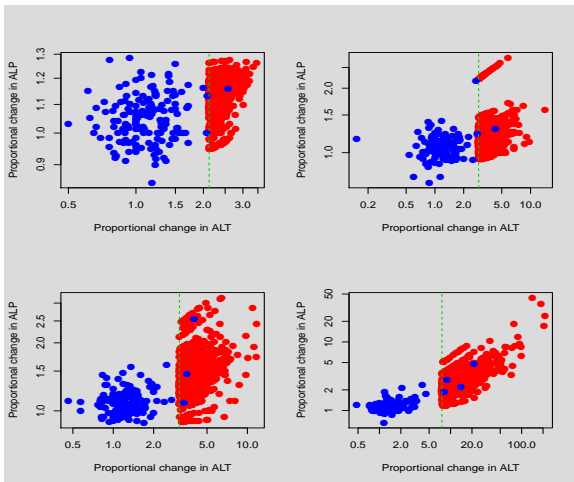
Observed ALP vs ALT



Predicted and observed TBL vs ALT



Predicted and observed ALP vs ALT



Some comments

- ▶ From experience, the relationship between ALT and ALP in this example is *not* typical
- ▶ The implication is that **some drugs associated with liver injury cause ALP to rise**
- ▶ Possibly, the lack of relationship between ALT and TBL is due to the short duration of treatment

Closing remarks

- ▶ Experience to date suggests
 - ▶ Extreme value modelling really can predict toxicities from early phase data
 - ▶ Approximately 50% of the time, outliers that appear worrying turn out to be **consistent with no treatment effect**
 - ▶ “Proceed with caution” rather than “clean bill of health”
 - ▶ Focus here has been on liver, but we've also analysed
 - ▶ Creatinine, LVEF, neutrophils, blood pressure

References I



U.S. Food and Drugs Administration.

Guidance for Industry Drug-Induced Liver Injury: Premarketing Clinical Evaluation.
2008.



Health Canada.

Guidance Document: Pre-market Evaluation of Hepatotoxicity in Health Products.
2012.



DCTD, NCI, NIH, and DHHS.

Cancer Therapy Evaluation Program, Common Terminology Criteria for Adverse Events, Version 3.0.
2003.



N. Kaplowitz.

Does elevated alkaline phosphatase exclude Hy's Law? No.
2010.