



Scatter plotting as a simple tool to analyse relative organ to body weight in toxicological bioassays

Felix M. Kluxen¹

Received: 23 May 2019 / Accepted: 26 June 2019 / Published online: 1 July 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

In toxicological bioassays, organ weight is often expressed as a ratio to body weight or another denominator to account for natural differences in animal sizes. However, the relationship of treatment-induced organ and body weight change is complicated and relative weights may accordingly confound a toxicological assessment. In addition, the statistical assessment of relative weights is challenging. The examples given in this document show that toxicological interpretation of organ weight data in relation to body weight can be vastly improved by simple bivariate scatter plotting. Conversely, plots of relative organ weight are of limited value and may lead to an incorrect interpretation of toxic effects when used in isolation. Scatter plots are useful for qualitative hazard characterization and to generate hypotheses. Bivariate summary statistics indicate effect levels and help to explore the actual correlation of organ to body weight.

Keywords Experimental toxicology · Regulatory toxicology · Exploratory data analysis · In vivo · Hazard characterization

Introduction

In toxicological bioassays, organ weight is often expressed as a ratio to body weight (Michael et al. 2007) or another denominator, to account for natural differences in animal sizes. There are, however, several issues with using relative organ weight data (Curran-Everett 2013), which have long been recognized (Angervall and Carlstrom 1963): the main problem is that an unbiased use of relative values is only possible if the regression line of nominator and denominator goes through the origin, i.e., the relationship of the ratio is constant (Bailey et al. 2004). If this is not the case, the use of relative values may result in a confounded toxicological assessment.

Organs' weight data often do not scale linearly or do not pass through the origin within species, because different organs might require a certain size range to function normally—independent of body weight. It has been shown in rats, which is the common test species used in toxicological assays, that some organs do not correlate well with bodyweight and some seem to actually correlate better with

brain weight (Bailey et al. 2004). Further, Trieb et al. (1976) showed that organ weights correlate better in an allometric power function with age, a method introduced by Huxley (1924). Hence, the required optimal relationship for relative values, as used in toxicology, is seldom achieved.

When a substance is tested in animals, it may affect a specific organ's weight, body weight and other aspects of the organs. When a substance interferes with feeding or/and potentially caloric efficiency, this may also modulate organ weight—it has been shown that organ weights change due to food restriction (Feron et al. 1973; Oishi et al. 1979; Takizawa 1978). Most treatment levels in regulatory toxicology assays are derived from preliminary dose range-finding experiments with body weight reduction as the primarily observed general toxic effect. While the exact mode of action of organ weight modulation does not necessarily affect the value of using organ weight as a gross hazard characteristic, this results in complicated relationships: the treatment may affect parameters at different levels, and the treatment may affect the correlation of organ to body weight, e.g., when organs of lighter animals are more affected by the treatment than those of heavier animals. Treatment-induced mean effects on organ and body weight itself affect mean relative organ weight values irrespective of affecting other regression parameters, see Table 1. For example, due to the dependence

✉ Felix M. Kluxen
Felix.Kluxen@adama.com

¹ ADAMA Deutschland GmbH, Cologne, Germany

Table 1 Type of effects on relative organ weight by treatment effects on organ and/or body weight

Effect on organ weight	Effect on body weight	Resulting relative organ weight	Example data
↑	↓	↑/↑↑	Figure 3
↑	↔	↑	Figure 5
↑	↑	↑/↔	Figure 1
↔	↓	↑	Figure 6
↔	↑	↓	— ^a
↓	↑	↓/↓↓	— ^a
↓	↓	↓/↔	Figure 4
↓	↔	↓	—

Arrows indicate an increase, decrease or no discernible effect. The effects ignore a potential change in correlation between the variables by treatment

^aA body weight increase is uncommon in preclinical toxicological bioassays

on both variables in the fraction, relative organ weights are affected by a specific effect on body weight. Hence, relative values can never be assessed in isolation and their use may be scrutinized based on this alone (Stevens 1976).

Most internationally accepted test guidelines, which are followed in toxicological studies for regulatory purposes, require the determination of organ weights relative to body and also brain weight, e.g., repeated-dose test guidelines from the Organisation for Economic Co-operation and Development (OECD). While the calculation of relative organ weights is recommended in toxicological studies (Sellers et al. 2007), there is no guidance on how to interpret potential differences in such comparisons. There is also no guidance for when treatment diametrically affects multiple organ weights or whether some or all organs should be excluded from the body weight value, which has been proposed relatively early in the use of relative organ weights (Cumming 1929), as terminal body weight is the sum of all organ weights, skin, fur, bones and carcass.

An appropriate statistical analysis of relative values and its interpretation is surprisingly challenging (Curran-Everett 2013). In practice, relative organ weights are often analysed by the same methods applied to the absolute values, e.g., by ANOVA/Dunnett's testing of the relative values themselves, which ignores any potential relationship of organ and body weight. A covariance approach may be more appropriate (Takizawa 1978), but may be hard to interpret (Hothorn 2016).

While the statistical interpretation of relative body weights can be difficult and misleading, a simple graphical bivariate scatter plotting may reveal an interpretable relationship that allows a toxicological assessment of organ to body weight or other ratio data by applying the principles of exploratory data analysis (Tukey 1977).

The current document describes the simple use of bivariate scatter plotting to analyse organ weight data in relation to body weight by examples and aims to make the method popular in the toxicological community. Scatter plotting can be performed by all software packages available to the researcher; here, the statistical software R (R Core Team 2017) and the ggplot2 package extension (Wickham 2016) were used. No animal studies were conducted for the current manuscript; all presented data come either from publically available sources (as referenced) or toxicological assays conducted for regulatory purposes (for the latter, not all tested groups are presented).

Using bivariate scatter plotting to analyse relative organ weights

The approach taken in this document is to assess organ weight relative to body weight by bivariate scatter plotting. This is an application of exploratory data analysis, which aims to present data in a way that creates insight and uncovers patterns—it does not concern formal inferential claims (Tukey 1977). It can be applied and understood with minimal statistical training.

Relative organ weights or other ratios or rates reduce two variables to a single one, which means that any information about their relationship is lost. By plotting the individual variables against each other—by bivariate scatter plots—and considering grouping information such as treatment levels by graphical methods, no information is lost and an informed toxicological qualitative hazard assessment is performed. There are several methods available to display grouping information in scatter plots: one can use different colours and shapes within the same plot or the dataset is spread over a plot array, called multi- or trellis plots (Cleveland 1985) or small multiples (Tuft 1990), which prevents overplotting.

Motivating example

Figure 1 depicts adrenal and body weight data presented and analysed in Angervall and Carlstrom (1963). The adrenal weight of two groups (A and B) is compared. Group B has an increased adrenal weight as compared to group A but also a higher body weight. While the absolute values are statistically significantly different, the organ weight ratios are not. The graphical presentation clearly shows that there are differences in adrenal weight (Fig. 1a) and body weight (Fig. 1c), but there is no difference in the relative organ weights (Fig. 1b). Based on only this information, one could conclude that the adrenal weights of group B are only higher, because the body weights of group B are higher than group A. However, Angervall and Carlstrom's comparative statistical analysis of “adjusted means” showed that “irrespective

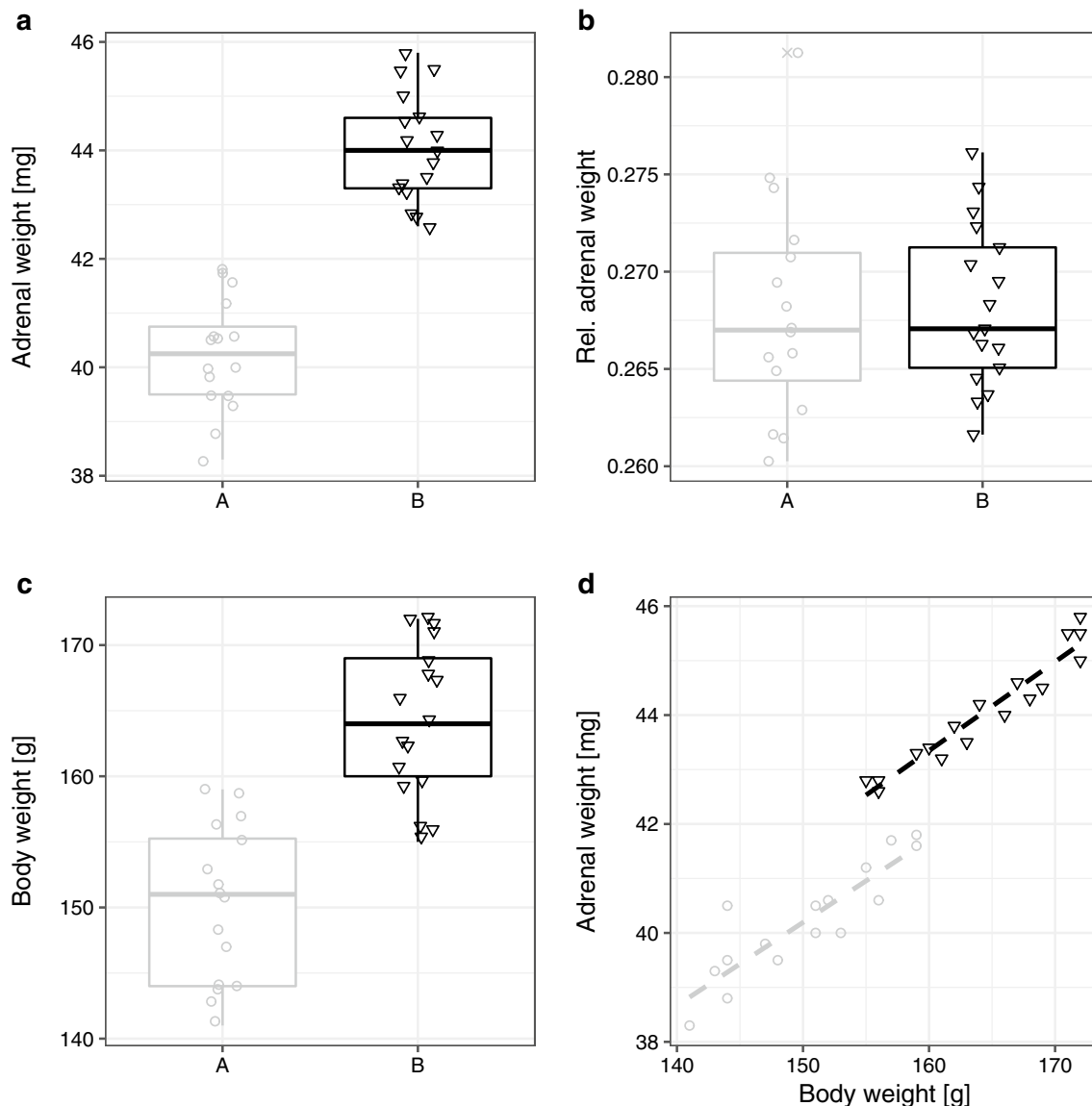


Fig. 1 Motivating example of a joint effect on organ and body weight. The adrenal and body weight data analysed by Angervall and Carlstrom (1963) are depicted. **a** Absolute adrenal weight, **b** relative adrenal weight, **c** body weight, and **d** organ to body weight overplotted by linear regression lines (dashed), coloured by group association. This

plot clearly shows that Group B (downward triangles) animals have higher adrenal weights by accounting for differences in body weight. The extreme values are depicted as an multiplication sign (x) next to the individual value in the boxplot for that group, because it exceeds 1.5 times the interquartile distance/the box (Tukey 1977)

of differences in body weight, the adrenal weights are significantly unequal”.

Scatter plotting leads to the same conclusion: Fig. 1d clearly shows, that group B animals have higher adrenal weights while accounting for differences in body weight. While animals with a higher body weight also have higher adrenal weights (the regression lines have a similar slope), the individual values are shifted towards the direction of higher adrenal weights, i.e., their regression line y-intercepts are different. If group B animals would have a higher adrenal weight only because of higher body weight, the regression

coefficients would be similar (group B regression line would be an extension of group A regression line).

The data from Angervall and Carlstrom (1963) are illustrative but seem less variable than usually observed in toxicological bioassays. Thus, further case studies, which exhibit more bivariate variation, are needed.

Visual guides and graphical summaries

While biological and toxicological graphs should include individual values for an unbiased assessment (Fosang and

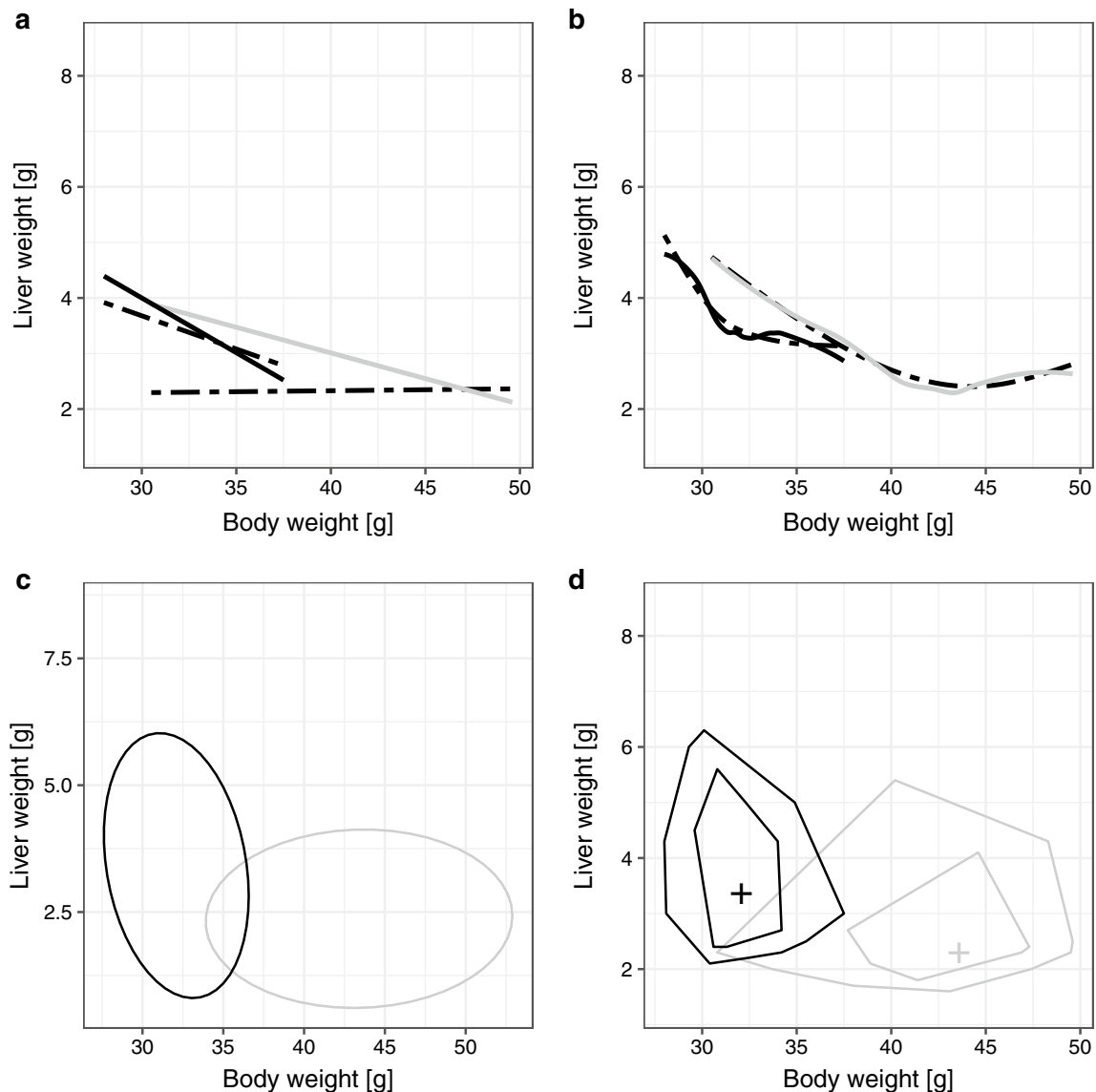


Fig. 2 Various summary methods for the individual values also depicted in Fig. 3 are shown. **a** Linear regression (solid lines) and robust linear regression (two dashed lines). **b** LOESS smoothing, i.e., locally weighted regression (solid lines) and natural cubic splines (two dashed lines—here very similar to loess). **c** Data ellipses. **d** Bag-plot, which is constructed based on “Tukey depth” and is the bivariate

approximation of the univariate boxplot and may be similarly interpreted. The central “plus” indicates the point of the highest Tukey depth, the inner polygon, the bag, consists of up to half of the data points. The outer polygon covers values that can be regarded as part of the distribution and values outside could be regarded as “outliers”

Colbran 2015; Nature Methods Editorial 2014; Pallmann and Hothorn 2016; Weissgerber et al. 2015), graphical summaries or guides help the viewer to detect patterns in scatter plots (Cleveland 1993). Various types of graphical guides are presented in Fig. 2 for the results of two treatment groups on liver and body weight. Common graphical guides are linear regression lines, but other methods are less common in toxicology such as robust regression (Marazzi and Joss 1993, both in Fig. 2a), scatter plot smoothers, i.e., locally weighted regression (also called locally estimated scatterplot smoothing, LOESS, Cleveland 1979), or smoothing splines

(Reinsch 1967, both in Fig. 2b). While those give a very good estimate of the dependence of “y” on “x”, they do not graphically summarize the bivariate distribution.

Individual responses in organ to body weight graphs can be summarized and their distribution visualized by data ellipses (Monette 1990, Fig. 2c), which have several useful properties (Friendly et al. 2013). They give a visual aid for grouping and variance and their slope is similar to the linear regression slope. Data ellipses are, however, biased by extreme values and assume bivariate normal distribution, hence robust alternatives could be developed. Friendly et al.

(2013) indicated methods using robust covariance estimates (Gnanadesikan and Kettenring 1972; Rousseeuw and Driessen 1999; Rousseeuw and Leroy 1987) in an early draft of the manuscript¹ that could be used to enhance ellipsoids. Another method was introduced by Rousseeuw et al. (1999), namely the “bagplot” (Fig. 2d), which is a bivariate generalization of the common boxplot. However, the practical application of such robust methods is currently restricted by limitations of the available software.² Data ellipses can be generated by multiple software packages and the approach may even be modified to allow inferential claims based on the assumed bivariate distribution (Guilbaud and Karlsson 2011; Thöni 1988; Wan et al. 2019). However, bagplots may be preferred for a more unbiased analysis—both ellipsoids (not adjusted for multiple comparisons) and bagplots are used as visual summaries in the manuscript. Univariate responses of absolute and relative organ weights and body weights in the examples are shown as individual values and superimposed boxplots (Tukey 1977), to allow an unbiased assessment of the results and not as mean and standard deviation as commonly used for toxicological bioassays.

Joint effects

Similarly to the data in Fig. 1, Figs. 3 and 4 also show joint effects on both organ and body weight that are commonly observed in toxicology.

Figure 3 shows the results of 18 months of treatment of male mice with a pesticide on body and liver weight. Both liver weight (Fig. 3a) and relative liver weight (Fig. 3b) are increased by treatment, while body weight is decreased (Fig. 3c). Figure 3d indicates a large bivariate variation in the responses and body weight retardation by the treatment, three values show an extreme response—two from the control and one in the treatment group, which are prone to confound common statistical analyses. While alternative robust statistical methods are available that are better suited to analyse data containing extreme values (Hothorn and Kluxen 2019), these are seldom used in the toxicological community. Figure 2a–d shows different summary methods for the individual values shown in Fig. 3d.

A correlation of body and organ weight is indicated only for the control group and only for robust linear regression (see Fig. 2b—two black dashed lines), which is less affected by extreme values than the linear regression. Organ and body weight seem inversely correlated in the treatment group. The graphical assessment in Figs. 2c, d and 3d highlights that the

treatment increases liver weight in addition to decreasing body weight.

Figure 4 shows the results of 13 weeks of gavage application of sodium dichromate dihydrate to female rats from the National Toxicology Program (NTP n.d.) as discussed in Hothorn (2016). Organ weight (Fig. 4a), relative organ weight (Fig. 4b) and body weight (Fig. 4c) are all decreased by the treatment. Accordingly, this is also seen in the organ vs body weight plot (Fig. 4d). The plot also shows a good correlation of organ and body weight within groups with similar slopes.

The pattern, i.e., ellipse shift towards the origin, indicates a general growth retardation induced by treatment. Further, the two high-dose ellipses have clearly lower liver weights, considering also lower body weights, and the highest dose is at the lower end of the body weight distribution, considering liver weight and one can accordingly formulate inferential hypotheses. The observation that ellipse sizes decrease with increasing dose helps to formulate a hypothesis about general growth retardation. Animals that presumably grow more than others in untreated conditions may be relatively more affected by treatment (refer to “Effect levels” for further assessment).

Isolated effects

The following cases describe isolated effect on either only organ weight (Fig. 5) or body weight (Fig. 6).

Figure 5 depicts the results of an uterotrophic assay according to OECD TG no. 440 (OECD 2007) with ethinyl estradiol by gavage application to female rats. Rats with removed ovaries (ovariectomized) are oestrogen depleted and have no functioning estrous cycle. As uterus size depends on oestrogen (and the cycle stage), depletion results in atrophy and an external treatment with an estrogenic agent results in increased uterus weight. This assay is used to identify estrogenic properties of chemicals.

There is a convincing induction of uterus weight upon oestrogen treatment (Fig. 5a), which is mirrored by the relative weight (Fig. 5b) but with no or only a weak effect on body weight (Fig. 5c). Absolute and relative uterus weight clearly show the same pattern and there is no gain of using relative weights, while there are issues with interpretation as outlined in the introduction. Conversely, the bivariate plot (Fig. 5d) is more informative than the relative uterus weight plot, because it shows the curious lack of correlation of organ and body weight, and does not obfuscate the effect on the absolute weights. An issue with the use of (normal) data ellipses becomes, however, apparent, as they are affected by extreme values.

Figure 6 shows the results of a 90-day rat study with dietary exposure to a pesticide. There is no effect on liver weight (Fig. 6a) and a pronounced effect on relative liver

¹ Friendly et al. (2013), <http://datavis.ca/papers/ellipses.pdf>.

² Bagplots can be generated using, e.g. R software and the `aplpack` package by H. P. Wolf or `ggplot2` together with some functions that are hosted on Ben Marwick’s GitHub repository.

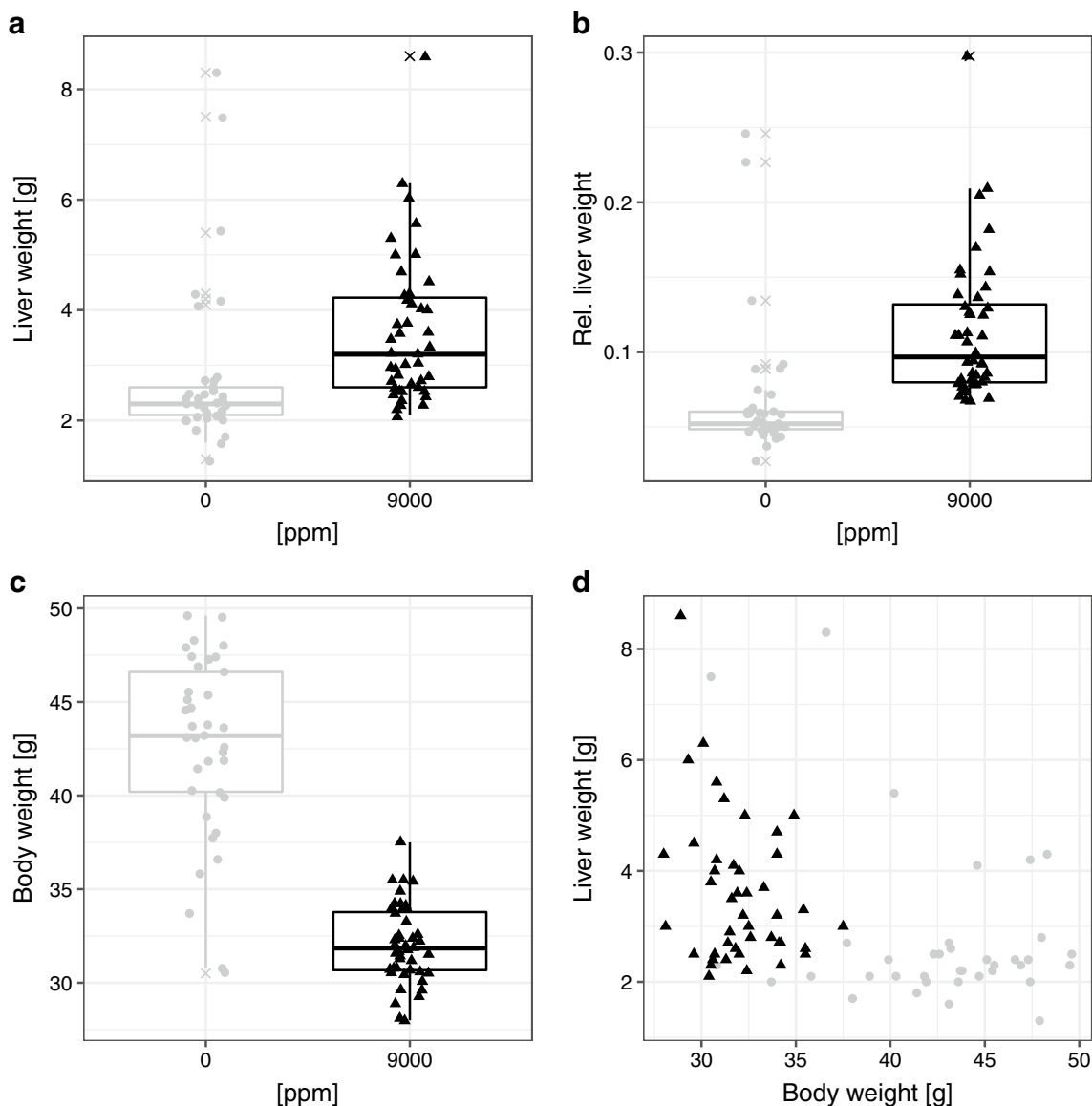


Fig. 3 Example of a joint effect on organ and body weight. The results of 18-month treatment of male mice with a pesticide on body and liver weight are shown. **a** Organ weight and **b** relative organ weight increased by treatment while **c** body weight decreased. Hence, the relative value might exacerbate the effect. **d** The liver weight

against body weight scatter plot allows a refined assessment, i.e., that treatment increases liver weight in addition to decreasing body weight. Several summary/visualization guide methods are shown in Fig. 2

weight (Fig. 6b), which seems to be a calculation artefact due to the effect on body weight (Fig. 6c). The bivariate plot confirms this (Fig. 6d) and indicates a decreased variance of the bivariate distribution, i.e., indications of growth retardation—a smaller ellipse for the high concentration as compared to the others.

Effect levels

Bivariate plotting can be used to estimate effect levels instead of using univariate relative organ weight data with

the associated issues (Curran-Everett 2013). Figure 7 shows the data already presented in Fig. 4, namely, liver weight data from a 13-week gavage application of sodium dichromate dihydrate to female rats from the National Toxicology Program. The dose levels progress over the plots from left to right and control and treatment levels are superimposed to allow a direct comparison.

The plot shows that the highest treatment groups behave clearly differently from the control. Hothorn (2016) presented multiplicity-adjusted p values for a bivariate Dunnett test for this dataset, which indicate a statistically significant effect

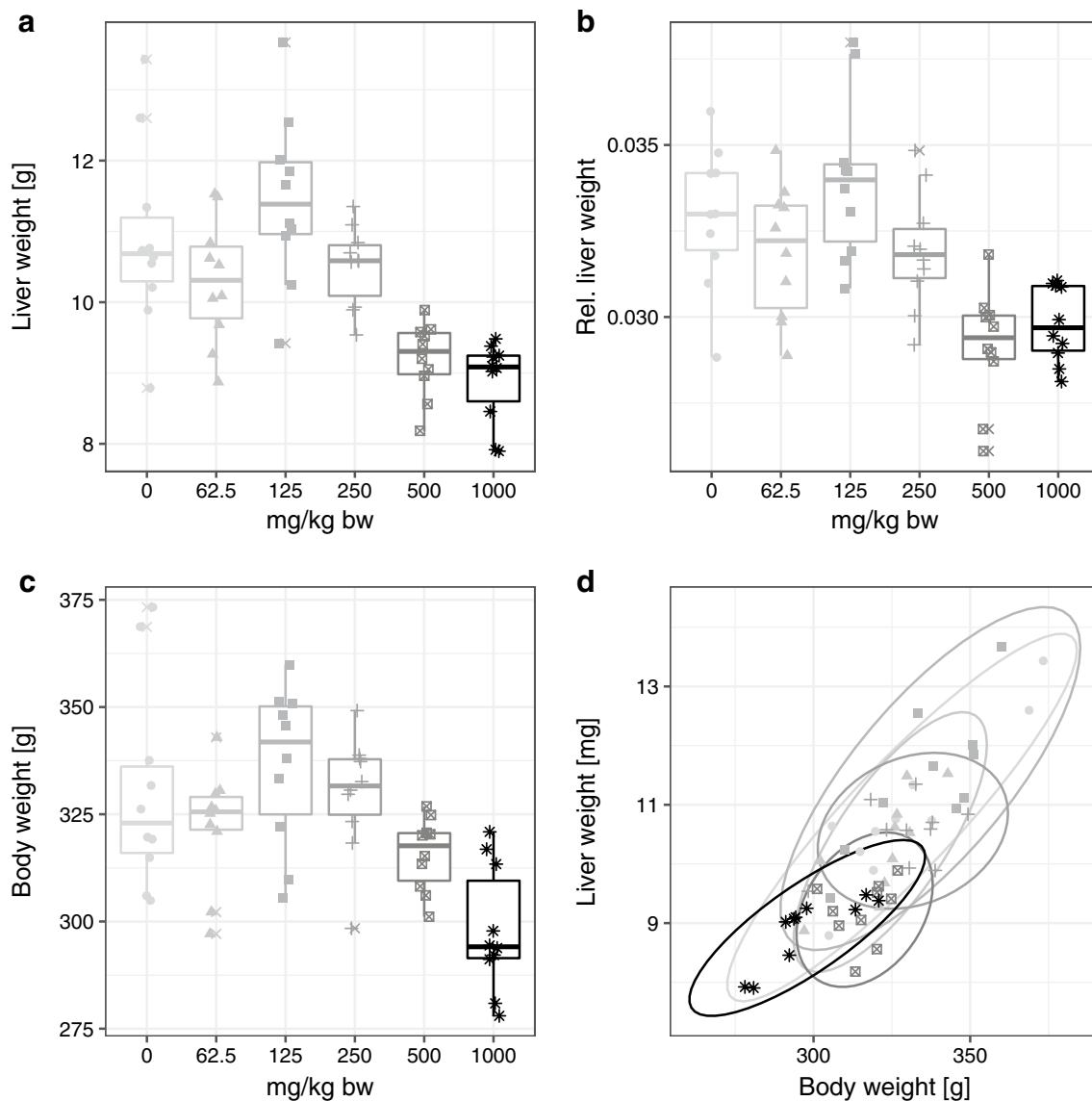


Fig. 4 Example of a joint effect on organ and body weight. the liver weight data from a 13-week gavage application of sodium dichromate dihydrate to female rats from the National Toxicology Program (NTP, n.d.) as discussed in Hothorn (2016) is depicted. **a** Organ weight, **b**

relative weight and **c** body weight all decrease with increasing dose. Hence, the relative value does not obscure the effect. **d** A good correlation of liver to body weight with similar linear regression coefficients, which is indicated by the ellipses is shown

on liver weight (at $p < 5\%$ alpha) for the 500 mg/kg bw/day group and a statistically significant effect for both liver and body weight at 1000 mg/kg bw/day. The graphical analysis supports the statistical assessment: the separation of the black treatment bagplot is more in the downward direction (liver weight) at 500 mg/kg bw/day but towards the origin at 1000 mg/kg bw/day.

Discussion

When the treatment affects both organ and body weight, a toxicological assessment based on relative values is prone to bias: relative values may obscure the actual relationship. This has been repeatedly discussed in literature (e.g.,

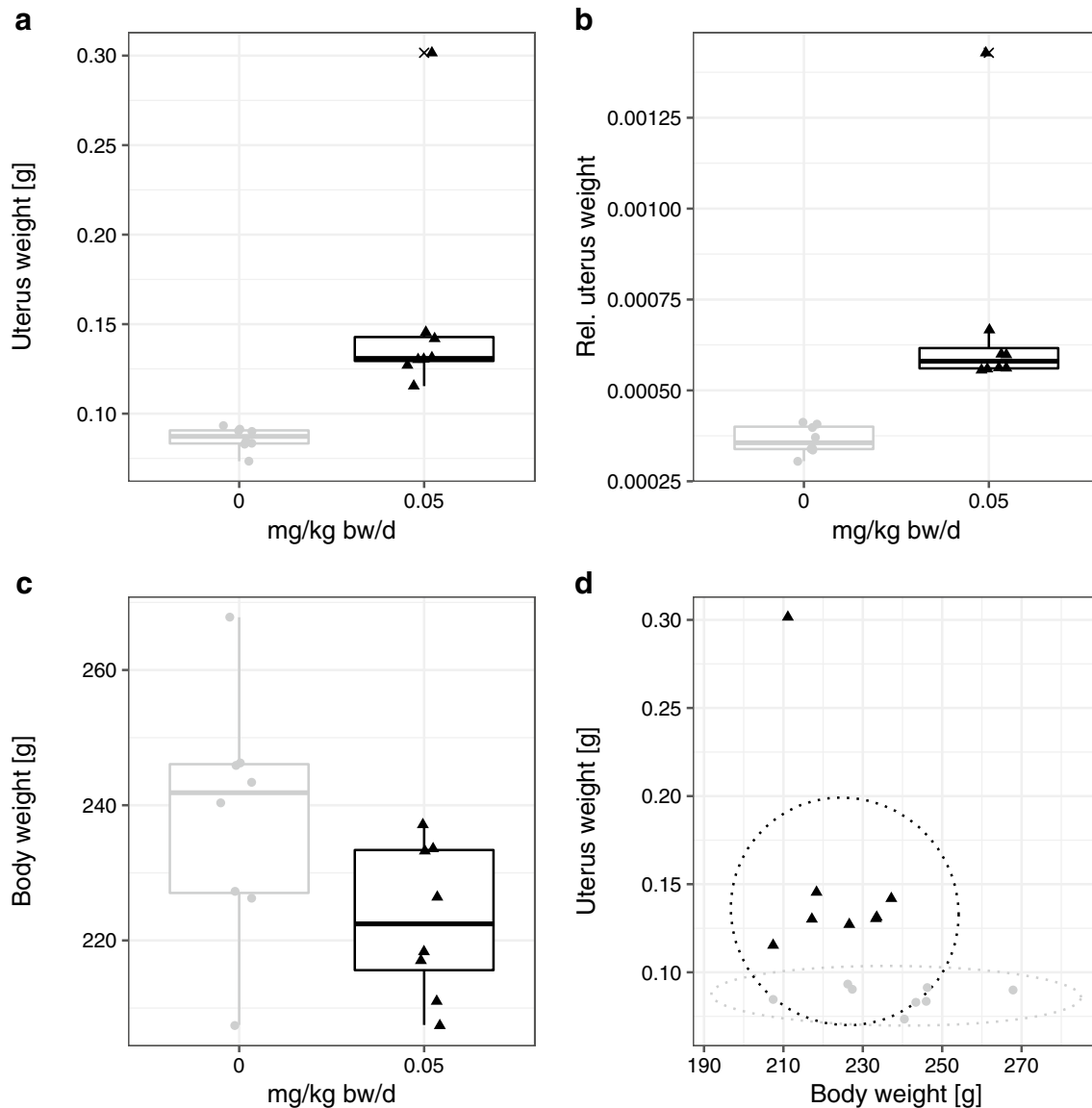


Fig. 5 Example of an organ weight-specific effect. The results of an uterotrophic assay, i.e., the gavage application of control or the oestrogen ethinyl estradiol (0.05 mg/kg bw/day) to female, oestrogen-depleted (ovariectomized) rats are depicted. **a** A convincing induction of uterus weight upon oestrogen treatment which is mirrored by the relative weight is shown (**b**) and no or only a weak effect on body weight (**c**). **d** The scatter plot shows a clear induction of uterus weight independent of body weight and that organ weight does not

correlate with body weight. The ellipses are not very useful here (and are accordingly depicted only with a dotted linetype), because they are sensitive to extreme values as seen for the ethinyl estradiol group. The extreme value is highlighted with a multiplication sign (×) next to the individual value triangle in the box plot for that group, because it exceeds 1.5 times the interquartile distance (box length) (Tukey 1977)

as reviewed in Bailey et al. 2004). Similarly, a statistical assessment of such values is problematic (Curran-Everett 2013; Hothorn 2016). Relative organ weights are used to account for different animal sizes and they may be preferred to decrease variation. Unfortunately, they do not (Stevens 1976)—and always have to be assessed together with the absolute organ and body weights.

The examples given in this manuscript demonstrate the use of scatter plotting to analyse organ weight depending on

body weight. It is an accessible tool that allows qualitative hazard characterization and the generation of hypotheses. It is thus of more value than plotting relative values and should be preferred when graphically presenting such data.

Scatter plots have long been used in science (Friendly and Denis 2005; Wainer 2013) and are very common in biological sciences. Their application for relative weight data is therefore expected to be understood by all peers. While their use in relation to body weight is common in the field of

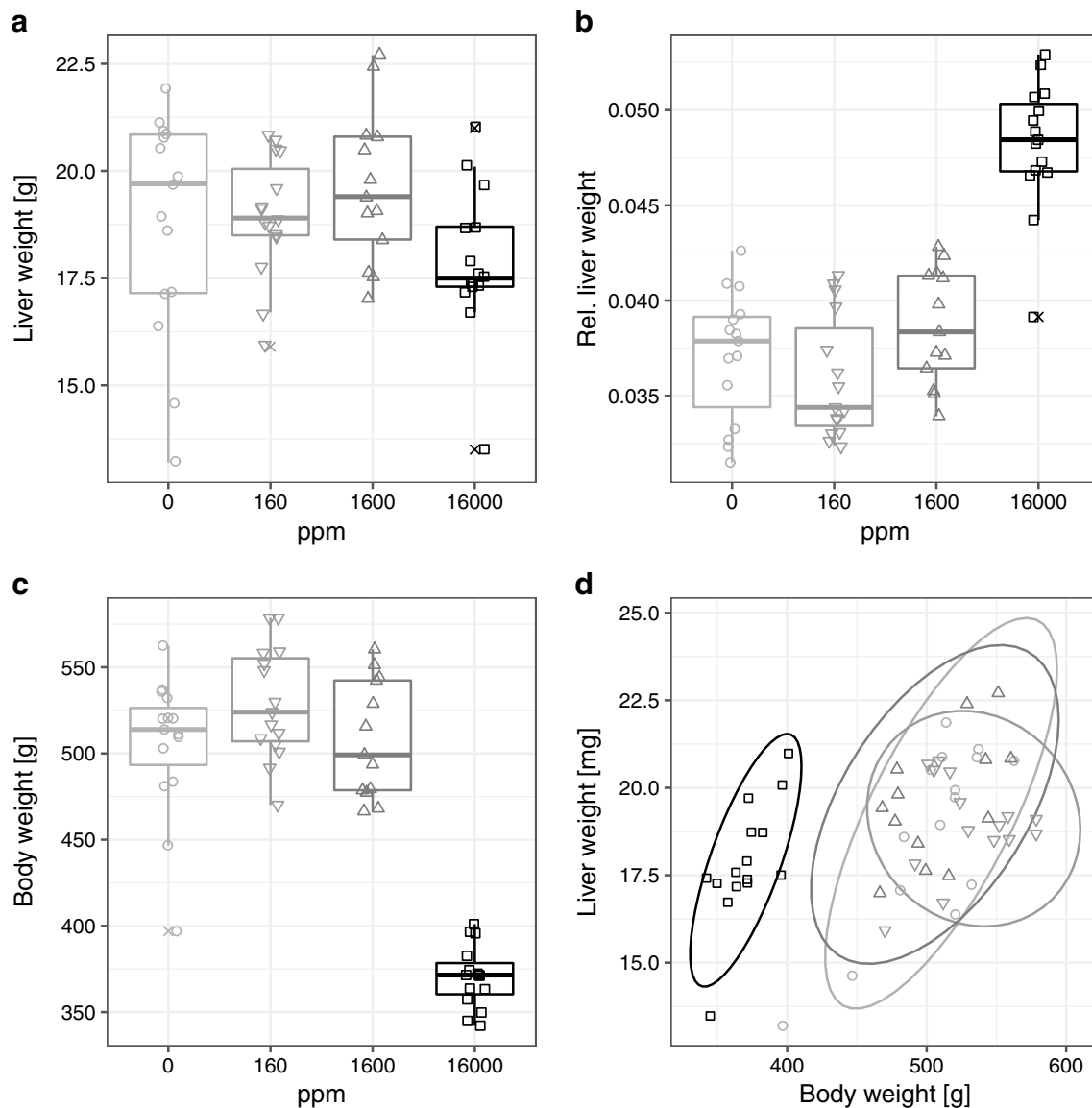


Fig. 6 Example of a body weight-specific effect. The results of dietary application of a pesticide to male rats for 90 days are depicted. **a** There is no effect on organ weight, **b** a strong effect on relative organ

weight and **c** a strong effect on body weight. **d** There is a good correlation between the weights and it is evident that only body weight is affected by the treatment, which drives the relative weight effect

allometry (Shingleton 2010), i.e., to compare animal strains (Anzai et al. 2017) or sexes (Heymsfield et al. 2007), it is uncommon in toxicology. While organ to body weight scatter plots were used to investigate correlation (Bailey et al. 2004) or to compare control populations in different laboratories (Weichenthal et al. 2010), they are not common while investigating the treatment effect. It is unclear why this is the case. Others observed the continuous use of uninformative bar plot means and standard deviation or error of the mean (Pallmann and Hothorn 2016; Weissgerber et al. 2015), while other summary methods such as the boxplot (Tukey 1977) have been available for decades and individual data points can be added to plots with all available graphing

software today. Hence, there seems to be a need to educate about graphical methods and the use of exploratory data analysis.

The statistical assessment and interpretation of relative weights are unfortunately not trivial. Angervall and Carlstrom (1963) described a method to compare differences in organ weight depending on body weight. Their comparative method of using adjusted means is, however, only applicable for two groups. Shirley (1977) describes the analysis of covariance for relative weights. While covariate analysis seems to be more appropriate than the use of the relative weights (Takizawa 1978), both methods can result in incorrect assessments, i.e., when body weight is affected

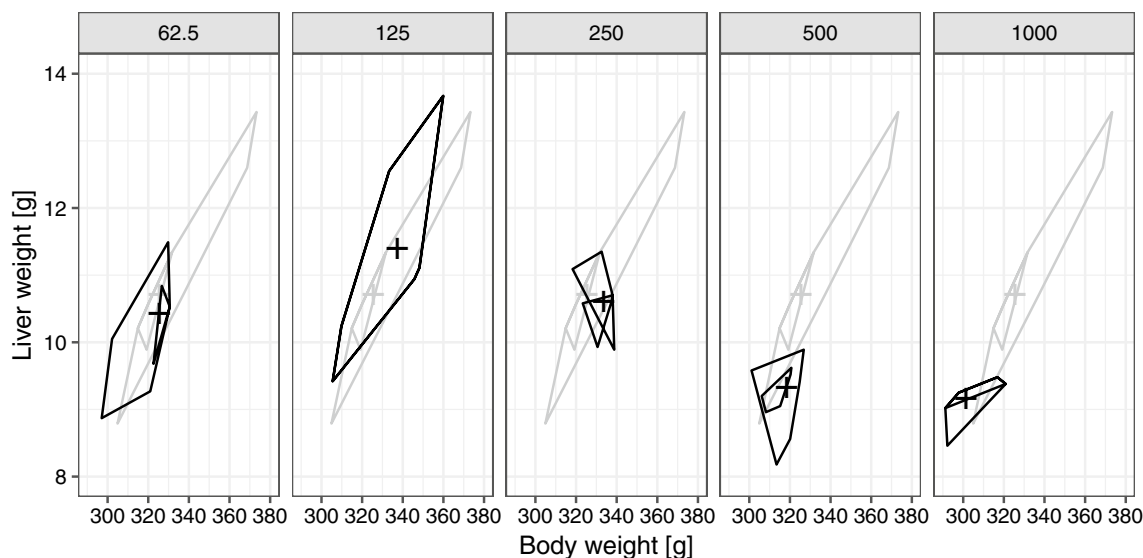


Fig. 7 Scatter plotting and superimposed bagplots can be used to explore effect levels. The data are already shown in Fig. 4, but here the treatment groups (in black) are shown over the range of a plot array together with the control bagplot (grey). The dose levels

are progressing from the left to the right plot in mg/kg bw/day and show liver weight against body weight, on y and x axis, respectively, as compared to the bagplot of the controls. The multiple plots allow pairwise comparisons

by treatment (Miller and Chapman 2001). Hothorn (2016) illustrates the issues of covariance analysis for the three most common organ to body weight relationships and proposes the application of a multivariate analysis (Andersen et al. 1999). This maintains alpha and allows the assessment of whether a weight effect occurs, at which dose on either organ or body weight. Overall, there seems to be no statistical ‘gold standard’ available. Further, there are no methods to derive benchmark doses for relative organ weights and Bayesian models are only in development.

Scatter plotting can be used to investigate the relationship of organ and body weight, however, it might be perceived to be inefficient to graphically investigate all collected organs weights in a toxicological bioassay. A research strategy might have to be developed on a case-by-cases basis.

If there is a change only in organ or body weight, one could make the case that the responses could be statistically compared in isolation. If there is a joint effect, a statistical bivariate analysis (Andersen et al. 1999; Hothorn 2016) can be applied. The statistical software R readily allows the application of multiple marginal models (Hothorn et al. 2008) in the multcomp package as described in Hothorn (2016) or Hothorn and Kluxen (2019).

Scatter plotting may be used to determine the strategy, help with the interpretation of the statistical result and can generate toxicological hypotheses. It may be extended by including multiplicity-adjusted summary statistics to make inferential claims and by plotting standardized values to compare the relative magnitude of observed effects (Festing 2014; Wan et al. 2019).

Conclusion

The toxicological interpretation of organ weight data in relation to body weight can be vastly improved by bivariate scatter plotting. Plots of relative organ weight are of limited value and may conversely lead to an incorrect interpretation of toxic effects when used in isolation. Scatter plots are useful for qualitative hazard characterization and help to generate hypotheses. Bivariate summary statistics indicate effect levels and help to explore the actual correlation of organ to body weight.

Acknowledgements I would like to thank the following people for educational discussions related to relative organ weights and associated issues: Dr. Stanley E. Lazic, AstraZeneca, for general discussions and the development of Bayesian models to assess toxic effects. Prof. Christian Ritz, University of Copenhagen, with regard to benchmark dose modelling and, Prof. Ludwig A. Hothorn for discussing different statistical methods to statistically compare relative organ weights.

Compliance with ethical standards

Conflict of interest The author declares that he has no conflict of interest.

References

- Andersen H, Larsen S, Spliid H, Christensen ND (1999) Multivariate statistical analysis of organ weights in toxicity studies. *Toxicology* 136(2):67–77. [https://doi.org/10.1016/S0300-483X\(99\)00056-6](https://doi.org/10.1016/S0300-483X(99)00056-6)

- Angervall L, Carlstrom E (1963) Theoretical criteria for the use of relative organ weights and similar ratios in biology. *J Theor Biol* 4(3):254–259
- Anzai H, Oishi K, Kumagai H, Hosoi E, Nakanishi Y, Hirooka H (2017) Interspecific comparison of allometry between body weight and chest girth in domestic bovinds. *Sci Rep* 7(1):4817. <https://doi.org/10.1038/s41598-017-04976-z>
- Bailey SA, Zidell RH, Perry RW (2004) Relationships between organ weight and body/brain weight in the rat: what is the best analytical endpoint? *Toxicol Pathol* 32(4):448–466. <https://doi.org/10.1080/01926230490465874>
- Cleveland WS (1979) Robust locally weighted regression and smoothing scatterplots. *J Am Stat Assoc* 74(368):829–836. <https://doi.org/10.1080/01621459.1979.10481038>
- Cleveland WS (1985) *The elements of graphing data*. Wadsworth Publ. Co., Belmont
- Cleveland WS (1993) *Visualizing data*. AT & T Bell Laboratories, Murray Hill, NJ, USA
- Cumming RR (1929) On the nature of hereditary size limitation. II. The growth of parts in relation to the whole. *J Exp Biol* 6(4):311–324
- Curran-Everett D (2013) Explorations in statistics: the analysis of ratios and normalized data. *Adv Physiol Educ* 37(3):213–219. <https://doi.org/10.1152/advan.00053.2013>
- Feron VJ, de Groot AP, Spanjers MT, Til HP (1973) An evaluation of the criterion “organ weight” under conditions of growth retardation. *Food Cosmet Toxicol* 11(1):85–94. [https://doi.org/10.1016/0015-6264\(73\)90064-3](https://doi.org/10.1016/0015-6264(73)90064-3)
- Festing MFW (2014) Extending the statistical analysis and graphical presentation of toxicity test results using standardized effect sizes. *Toxicol Pathol* 42(8):1238–1249. <https://doi.org/10.1177/0192623313517771>
- Fosang AJ, Colbran RJ (2015) Transparency is the key to quality. *J Biol Chem* 290(50):29692–29694. <https://doi.org/10.1074/jbc.E115.000002>
- Friendly M, Denis D (2005) The early origins and development of the scatterplot. *J Hist Behav Sci* 41(2):103–130
- Friendly M, Monette G, Fox J (2013) Elliptical insights: understanding statistical methods through elliptical geometry. *Stat Sci* 28(1):1–39. <https://doi.org/10.1214/12-STS402>
- Gnanadesikan R, Kettenring JR (1972) Robust estimates, residuals, and outlier detection with multiresponse data. *Biometrics* 28:81–124
- Guilbaud O, Karlsson P (2011) Confidence regions for Bonferroni-based closed tests extended to more general closed tests. *J Biopharm Stat* 21(4):682–707. <https://doi.org/10.1080/10543406.2011.551331>
- Heymsfield SB, Gallagher D, Mayer L, Beetsch J, Pietrobelli A (2007) Scaling of human body composition to stature: new insights into body mass index. *Am J Clin Nutr* 86(1):82–91. <https://doi.org/10.1093/ajcn/86.1.82>
- Hothorn LA (2016) *Statistics in toxicology using R*. CRC Press, Boca Raton
- Hothorn LA, Kluxen FM (2019) Robust multiple comparisons against a control group with application in toxicology. [arXiv:1905.01838](https://arxiv.org/abs/1905.01838) [stat.AP]
- Hothorn T, Bretz F, Westfall P (2008) Simultaneous inference in general parametric models. *Biom J* 50(3):346–363. <https://doi.org/10.1002/bimj.200810425>
- Huxley JS (1924) Constant differential growth-ratios and their significance. *Nature* 114:895–896
- Marazzi A, Joss J (1993) Algorithms, routines, and S functions for robust statistics: the FORTRAN library ROBETH with an interface to S-PLUS. Wadsworth Publ. Co., Belmont
- Michael B, Yano B, Sellers RS et al (2007) Evaluation of organ weights for rodent and non-rodent toxicity studies: a review of regulatory guidelines and a survey of current practices. *Toxicol Pathol* 35(5):742–750. <https://doi.org/10.1080/01926230701595292>
- Miller GA, Chapman JP (2001) Misunderstanding analysis of covariance. *J Abnorm Psychol* 110(1):40–48. <https://doi.org/10.1037/0021-843X.110.1.40>
- Monette G (1990) Geometry of multiple regression and interactive 3-D graphics. In: Fox J, Long JS (eds) *Modern methods of data analysis*. Sage, Beverly Hills, pp 209–256
- National Toxicology Program (n.d.) 13 weeks gavage Study on female F344 rats administered with sodium dichromate dihydrate (VI) (CASRN: 7789-12-0, Study number: C20114. TDMS number: 2011402) Technical report
- Nature Methods Editorial (2014) Kick the bar chart habit. *Nat Methods* 11:113. <https://doi.org/10.1038/nmeth.2837>
- OECD (2007) Test no. 440: uterotrophic bioassay in rodents. OECD Publishing, Paris. <https://doi.org/10.1787/9789264067417-en>
- Oishi S, Oishi H, Hiraga K (1979) The effect of food restriction for 4 weeks on common toxicity parameters in male rats. *Toxicol Appl Pharmacol* 47(1):15–22. [https://doi.org/10.1016/0041-008X\(79\)90066-8](https://doi.org/10.1016/0041-008X(79)90066-8)
- Pallmann P, Hothorn LA (2016) Boxplots for grouped and clustered data in toxicology. *Arch Toxicol* 90(7):1631–1638. <https://doi.org/10.1007/s00204-015-1608-4>
- R Core Team (2017) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna
- Reinsch CH (1967) Smoothing by spline functions. *Numer Math* 10(3):177–183. <https://doi.org/10.1007/bf02162161>
- Rousseeuw PJ, Driessen KV (1999) A fast algorithm for the minimum covariance determinant estimator. *Technometrics* 41(3):212–223. <https://doi.org/10.1080/00401706.1999.10485670>
- Rousseeuw PJ, Leroy AM (1987) *Robust regression and outlier detection*. Wiley, New York
- Rousseeuw P, Ruts I, Tukey JW (1999) The bagplot: a bivariate boxplot. *Am Stat* 53:382
- Sellers RS, Mortan D, Michael B et al (2007) Society of toxicologic pathology position paper: organ weight recommendations for toxicology studies. *Toxicol Pathol* 35(5):751–755. <https://doi.org/10.1080/01926230701595300>
- Shingleton A (2010) Allometry: the study of biological scaling. *Nat Educ Knowl* 3(10):2
- Shirley E (1977) The analysis of organ weight data. *Toxicology* 8(1):13–22
- Stevens MT (1976) The value of relative organ weights. *Toxicology* 5(3):311–318
- Takizawa T (1978) An unbiased comparison of organ weights when an inequality in body weight exists. *Toxicology* 9(4):353–360. [https://doi.org/10.1016/0300-483X\(78\)90018-5](https://doi.org/10.1016/0300-483X(78)90018-5)
- Thöni H (1988) Multiple comparisons and conditional joint confidence regions. In: Bauer P, Hommel G, Sonnemann E (eds) *Multiple Hypothesenprüfung / Multiple Hypotheses Testing: Symposium*, 6. und 7. November 1987. Springer, Berlin, Heidelberg, pp 190–206. https://doi.org/10.1007/978-3-642-52307-6_16
- Trieb G, Pappritz G, Lützen L (1976) Allometric analysis of organ weights. I. Rats. *Toxicol Appl Pharmacol* 35(3):531–542. [https://doi.org/10.1016/0041-008X\(76\)90076-4](https://doi.org/10.1016/0041-008X(76)90076-4)
- Tufte E (1990) *Envisioning information*. Graphics Press, Cheshire, CT, USA
- Tukey JW (1977) *Exploratory data analysis*. Addison-Wesley Pub. Co., Reading
- Wainer H (2013) *Graphic discovery: a trout in the milk and other visual adventures*. Princeton University Press, Princeton
- Wan F, Kunz CU, Jaki TF (2019) Confidence regions for treatment effects in subgroups in biomarker stratified designs. *Biom J* 61(1):27–39. <https://doi.org/10.1002/bimj.201700303>
- Weichenthal S, Hancock S, Raffaele K (2010) Statistical power in the analyses of brain weight measures in pesticide neurotoxicity testing and the relationship between brain and body weight. *Regul*

Toxicol Pharmacol 57(2):235–240. <https://doi.org/10.1016/j.yrtph.2010.03.001>

Weissgerber TL, Milic NM, Winham SJ, Garovic VD (2015) Beyond bar and line graphs: time for a new data presentation paradigm. PLoS Biol 13(4):e1002128. <https://doi.org/10.1371/journal.pbio.1002128>

Wickham H (2016) ggplot2—Elegant graphics for data analysis, 2nd edn. Springer International Publishing, Cham

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.