Pseudo-data generation allows the statistical re-evaluation of toxicological bioassays based on summary statistics

Ludwig A. Hothorn, Felix M. Kluxen, Mario Hasler
Leibniz University Hannover, ADAMA Deutschland GmbH,
Christian-Albrechts-University of Kiel, Germany

October 18, 2019

Abstract

Sometimes a re-analysis of toxicological data is useful. However, this usually requires the availability of the original data and in many cases only summary data are available in the publications. Here the generation of pseudo-data under certain assumptions using extension packages in the open-source project R on statistical computing is shown. Several case studies are used to illustrate the applicability in regulatory toxicology.

1 Introduction

Results in publications on toxicological bioassays are usually presented as summary statistics; often in bar chart plots or table format, showing means, standard deviations and sample sizes. This practice has been repeatedly criticized (6; 2; 21) because potential sub-grouping in the data is hidden and the validity of the performed statistical analysis cannot be peer-reviewed. In internationally accepted test guidelines for regulatory bioassays, the reporting of individual data is mandatory, but often not publicly available, because they may support commercial interests. However, summary data usually are available, e.g. from the European Chemicals Agency (ECHA) or the European Food Safety Authority (EFSA), depending on the regulatory field.

When data are presented as means and standard deviations, the authors stating such values implicitly assume normal distribution of the reported data, irrespective of the actual underlying empirical distribution of the data. Therefore, it is possible to generate individual observations, here denoted as pseudo-data, from these summary statistics using a simulation algorithm.

Pseudo-data can be used for peer-review, re-analysis by different or novel methods, educational purposes and for incorporating multiple data sets in an analysis, when only summary statistics are available. While the original data are obviously more valuable, the pseudo-data approach can be a compromise; e.g. when individual data is hard to obtain or not electronically available - the latter is sometimes the case for historical control data.

We demonstrate the use of the simulation algorithm in the function `ermvnorm` within the *SimComp* package, available for the open-source statistical software R. This function is based on the function `rmvnorm` of the package *mvtnorm*. This method differs from other random data generation algorithms in that a multivariate normal distribution with an exact nominal mean vector, exact variance vector and approximate correlation matrix (for multiple endpoints or repeated measures) is calculated. (Note that random numbers with exact mean and standard deviation cannot be used for simulations concerning type I error or power of statistical tests.)

Multiple case studies are presented and an appropriate re-analysis is demonstrated: i) repeatability by a second assay, ii) fold change effect size, iii) taking variance heterogeneity into account, iv) considering dose as quantitative covariate and/or qualitative factor, v) user-defined contrasts, and vi) multiple endpoints. In a seventh case study the limitation of the approach is shown because the normal distribution assumption failed.

The presented method can be easily applied by statisticians and toxicologists by using the example R-code provided within this manuscript.

2 Generating random numbers with exact mean and standard deviation

First, summary statistics are selected, e.g. from a publication table for 7 antioxidant indices in epididymides of mice exposed to 0 (control), 25 (dose1), 50 (dose2) and 100 (dose3) mg/L sodium fluoride as shown by the following table from (19). Second, sample sizes (n), arithmetic means (mean) and standard errors (sd) for a particular endpoint, for example the low total antioxidant capacity (T-AOC), are simply transferred to the following R code:

Groups	Control	25 mg/L NaF	50 mg/L NaF	100 mg/L NaF
ROS (fluorescence intensity/mgprot)	1640 ± 236	1670 ± 215	1713 ± 306	2359 ± 437
T-AOC (U/mgprot)	6.62 ± 2.01	2.84 ± 0.99	3.30 ± 0.23	1.98 ± 0.44*
MDA (nmol/mgprot)	1.91 ± 0.59	3.61 ± 0.55	4.51 ± 0.59	7.46 ± 1.33**
GSH (μmol/L)	17.56 ± 1.19	9.37 ± 0.73**	8.60 ± 0.29**	7.79 ± 0.33**
GPx (U/mgprot)	55.40 ± 7.10	45.18 ± 6.44	40.65 ± 4.05	33.07 ± 4.32*
GR (U/mgprot)	11.52 ± 1.71	8.90 ± 1.22	8.18 ± 0.58	6.36 ± 0.94*
GST (U/mgprot)	34.56 ± 5.73	21.82 ± 2.88	21.65 ± 0.87	18.29 ± 2.00*

Figure 1: Table: Summary data table in Sun et al., 2018

```
library(SimComp) # loading the selected CRAN package
set.seed(127711) # selecting a seed allows reproducibility
control <- ermvnorm(n=6 ,mean=6.62, sd=2.01*sqrt(6))
dose1 <- ermvnorm(n=6 ,mean=2.84, sd=0.99*sqrt(6))
dose2 <- ermvnorm(n=6 ,mean=3.30, sd=0.23*sqrt(6))
dose3 <- ermvnorm(n=6 ,mean=1.98, sd=0.44*sqrt(6))
dose <- factor(rep(c("0","25","50","100"), c(6,6,6,6)))
TAOC <- c(control, dose1, dose2, dose3)
datTA <- data.frame(TAOC, dose)
```

The object `TAOC` contains the pseudo-data with the qualitative factor `dose` and can be used for re-analysis. These data can be easily visualized by modified box plots using the R-package `toxbox` (13)
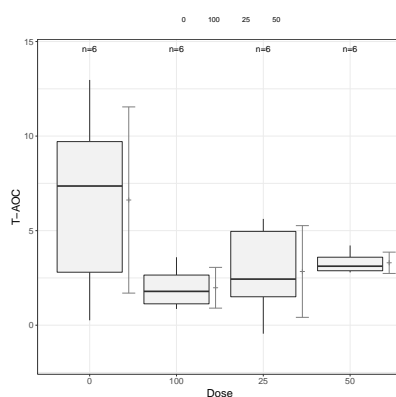

Figure 2: Box plots of T-AOC pseudo-data.

An extension of this example considering multiple correlated endpoints and repeated measures for assumed or random correlations is available and is demonstrated in the sixth case study below.

The simple idea behind the function `ermvnorm` is to apply the function `rmvnorm` of the package *mvtnorm*. This function generates $n - 2$ random vectors from a k-dimensional multivariate normal distribution whose mean vector is given by `mean` and corresponding vector of standard deviations (SD) by `sd`. If only one endpoint is of interest, dimension k equals 1. For more details, see (14). These $n - 2$ random vectors do not yet have the indented exact mean and SD. However, based on the first $n - 2$ vectors, a system of equations is formed by the usual formulas of sample mean and sample variance. The remaining 2 vectors are obtained as solution so that the specified values for the mean vector `mean` and for the vector of standard deviations `sd` are achieved. Depending on the first $n - 2$ random vectors, the underlying system of equations may have no solution. In this case, `ermvnorm` creates a new set of $n - 2$ random vectors until a valid pseudo-data set is obtained or until the maximum number of tries is reached.

3 Case studies

The following case studies show the applicability of the method for re-analysis for several common statistical scenarios encountered in toxicology.

3.1 Demonstrating repeatability

Demonstrating repeatability of a previously observed effect appears relatively simple, e.g. by a second assay under comparable conditions and seems necessary in the current reproducibility crisis. (1) argue that this might be due insufficient statistical expertise, i.e. sampling effects. Usually, data of a new assay will be compared with data from a previous assay in the literature. In this example, data for the first and the second assay were provided within the same project for the hematology endpoint mean corpuscular volume (MCV) for mice that were treated with five concentrations of sodium dichromate dihydrate (18). Summary statistics were extracted and pseudo-data generated. The boxplots in Figure 3 reveal a rather similar pattern for both studies (parallel vertical dotted lines: left study 1, right study 2) with a general tendency to lower values in the second study.
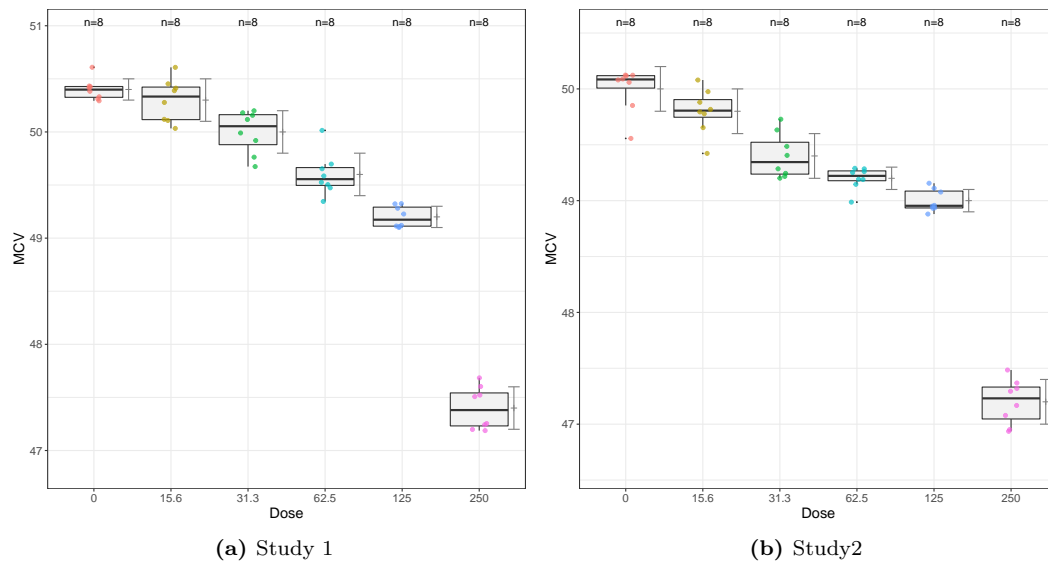
2

Figure 3: Box-plots MCV pseudo data

Table 1 shows similar effect sizes (Effect size 1, Effect size 2), i.e. the mean difference $\mu_i - \mu_0$, and similar tiny p-values (p-value 1, p-value 2) for almost all comparisons against control for both study 1 and study 2, which is a convincing empirical criterion for repeatability. Irrespective of the observed additive shift, the patterns of effect sizes and p-values are rather similar in both studies. Alternatively, one can use a more complex model, a mixed effects model with a random factor between both studies. Again, the effect size (Effect size mixed) and the p-value (p-value mixed) are quite similar.

Table 1: Effect sizes and p-values for Dunnett tests for MCV pseudo data for assay 1 and assay 2 and their joint effect size

	Comparisons	Effect size 1	p-value 1	Effect size 2	p-value 2	Effect size mixed	p-value mixed
1	15.6 - 0	-1.2	0.66860	-2.3	0.09902	-2.3	0.09306
2	31.3 - 0	-4.6	0.00015	-6.9	0.00000	-7.7	0.00000
3	62.5 - 0	-9.2	0.00000	-9.2	0.00000	-12.3	0.00000
4	125 - 0	-13.9	0.00000	-11.5	0.00000	-16.9	0.00000
5	250 - 0	-34.6	0.00000	-32.3	0.00000	-44.4	0.00000

A third and quite different approach is the claim of no interactions as a criterion for repeatability using an interaction contrast test (10). The dose-by-study interaction contrasts are estimated in Table 2 revealing quite different slopes for $(\mu_{31}^1 - \mu_0^1) - \mu_{31}^2 - \mu_0^2)$ and $(\mu_{250}^1 - \mu_0^1) - \mu_{250}^2 - \mu_0^2)$, i.e. a substantial higher sensitivity of study 2 with respect to study 1. Hereby it is difficult to conclude on similarity of the two studies, as this is injured evenly with two dose comparisons substantially.

Table 2: Effect sizes and p-values for interaction contrasts for MCV data

	Interaction contrast	Effect size	p-value
1	((15.6 - 0):2) - ((15.6 - 0):1)	-0.30	0.017
2	((31.3 - 0):2) - ((31.3 - 0):1)	-1.70	0.000
3	((62.5 - 0):2) - ((62.5 - 0):1)	-0.10	0.781
4	((125 - 0):2) - ((125 - 0):1)	-0.10	0.781
5	((250 - 0):2) - ((250 - 0):1)	-1.70	0.000

3.2 Fold change as effect size

The majority of statistical significance tests use (implicitly) the difference of means as effect size. However, the ratio of means, denoted as fold change, is used in toxicology for specific interpretations, such as the 2-fold rule in the Ames assay (4). Unfortunately, this multiplicative effect size is sometimes used together with t- or Wilcoxon test for an additive effect size, e.g. for urine excretion in rats treated with seven doses of hexachloro-butadiene (20)(in their Table 5, see the simulated data in Table 3). The common approach to switch into the multiplicative model is the log-transformation (which requires the validity of a log-normal distribution, see e.g. (16)), but for normal distributed endpoints a modified Dunnett or Williams test for ratios-to-control is available (9) as an alternative.

The actual data from (20) is reproduced in Table 3 and the generated pseudo-data depicted in Figure 4.

3

	UE	dose
1	10.02	0
2	7.47	0
3	10.41	0
4	9.65	5
5	7.28	5
6	10.07	5
7	12.51	10
8	11.71	10
9	13.88	10
10	18.74	15
11	17.28	15
12	18.58	15
13	24.70	20
14	16.11	20
15	25.79	20
16	70.57	30
17	26.29	30
18	40.54	30
19	16.23	45
20	14.33	45
21	61.84	45
22	21.64	90
23	20.28	90
24	72.39	90

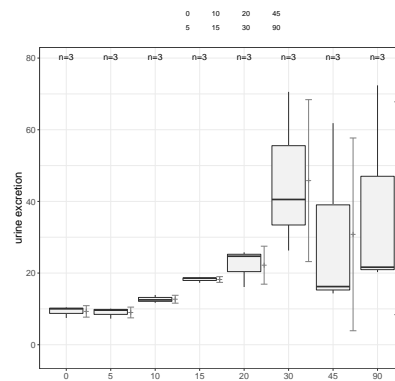**Table 3:** Simulated urine excretion data

Figure 4: Box plot urine excretion pseudo data.

In this balanced design with extreme small sample sizes, an extreme variance heterogeneity can be seen for higher doses in Figure 4. A modified Dunnett-type test for ratios-to-control allowing heterogeneous variance is used therefore (7).

	Fold change	p-value
5/0	0.97	1.00
10/0	1.37	0.20
15/0	1.96	0.04
20/0	2.39	0.17
30/0	4.92	0.31
45/0	3.31	0.71
90/0	4.10	0.59

Table 4: Ratio-to-control inference for $n = 3$.

The fold changes are well reproduced from Tab. 5 of (20), but the Dunnett test for ratio-to-control (assuming normal distributed errors with any variance) reveals a weak significant increase for the 15 mg dose only (adjusted p-value) – primarily induced by this extreme small variance, i.e. the three similar values such as $(18.7, 17.3, 18.6)$ by a value range from about 7.2 to 72.4. There is no biological sound increase at any dose – induced by $n_i = 3$ and the extreme variances at high doses. The conclusion by the authors ”*Urinary KIM-1 excretion was increased at doses between 10 and 30mgkg, but owing to a high group variance (i.e. large SDs), the increases of 3.31- and 4.10-fold at, respectively, 45 and 90mgkg HCBD, compared with the control group, were not statistically significant*” can not be reproduced. This example shows very nicely what the p-values would have looked like if slightly higher sample sizes had been used, e.g. $n_i = 6$, just running the simulation code with $n = 6$, assuming that the higher sample size would have the same means and SD.

The well-known fact that the sample size has a substantial influence on the p-value becomes very clear, which is particularly important in toxicology.

3.3 Taking variance heterogeneity into account

The common assumption of homogeneous variances in dose-responses relationships may be sometimes unrealistic. Corresponding tests tend to be liberal when higher variances occur in higher doses, particularly when sample sizes are smaller in the dose groups (vice versa). In these cases, a modified Dunnett-type test allowing heterogeneous variance

	Fold change	p-value
5/0	0.97	1.000
10/0	1.37	0.015
15/0	1.96	0.000
20/0	2.39	0.011
30/0	4.92	0.051
45/0	3.31	0.405
90/0	4.10	0.260

Table 5: Ratio-to-control inference for $n = 6$.

should be used (7). As an example we use summary data of pregnant mice which were exposed to five phthalates and their mixture. In their Table 2 (3), the body weight mean, SD and n for each treatment group is reported together with stars for Dunnett's post hoc test. From the simulated data in Figure 5 an unbalanced design with heterogeneous variance can be seen.
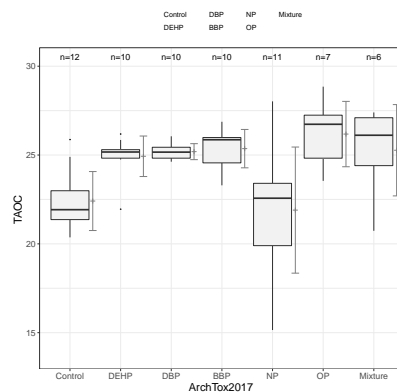

Figure 5: Box plot body weights.

Table 6: Modified Dunnett test allowing heterogeneous variances

	Comparison	Dunnett	Dunnett var het
1	DEHP - Control	0.024	0.001
2	DBP - Control	0.010	0.000
3	BBP - Control	0.006	0.000
4	NP - Control	0.981	0.997
5	OP - Control	0.001	0.001
6	Mixture - Control	0.029	0.118

Consequently, the re-analysis with the Dunnett test assuming homogeneous variance reproduced the significance signs. But taking variance heterogeneity into account, there is no significant body weight change in the mixture, but more significant changes in the phthalate groups. This is not supported by the conclusion by the authors *Data showed that, except for NP, all mice exposed to EDCs, single or mixed, presented higher BW than control mice.*

3.4   Considering dose as quantitative covariate and/or qualitative factor

Trend tests are classified into those that accept the dose as a factor, i.e. qualitative, and those that accept the dose as a covariate, i.e. quantitative. Both concepts have advantages and disadvantages, but they can even be combined (15). As an example the clinical chemistry endpoints of a feeding study with GMO-corn for the female rats, where the endpoint CHO was selected (Table 4) (5).
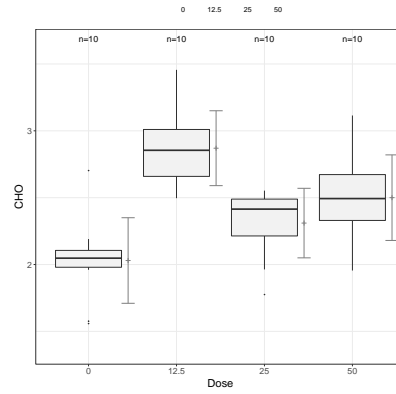
5

Figure 6: Boxplot CHO endpoint in female rats.

Table 7: Tukey and Williams Trend Test

Dose metameter	Test statistic	p-value
Doseari: Doseari	1.2759	0.2471
Doseord: Doseord	1.4505	0.1898
Doselog: Doselog	-2.4577	0.9999
Dosearilog: Dosearilog	1.4505	0.1897
Dosetreat: C 1	3.5488	0.0007
Dosetreat: C 2	3.2176	0.0023
Dosetreat: C 3	4.7718	0.0000

Assuming the concentration as a quantitative covariate, no significant increasing trend reveals. However, taking additionally the Williams contrast with dose as factor into account, a clear significant plateau effect (contrast C3) follows. This finding is more precise than the significance signs for the 12.5 and 50 group in Table 4 of (5).

3.5   User-defined comparisons

Sometimes in trend analysis, additional comparisons than against control are of interest, e.g. control versus high dose (11). As an example, the endpoint VAP (average path velocity) in the gestational exposure to Bisphenol-A affects the spermatozoa in adult mice (12).
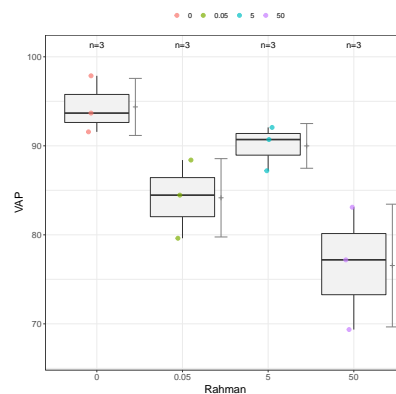

Figure 7: Boxplot VAP.

Table 8: User-defined comparisons

	Comparison	Test statistic	p-value
Trend with arithmetic scores	Doseari: Doseari	-3.346	0.001
Trend with ordinal scores	Doseord: Doseord	-3.067	0.002
Trend with logarithmic scores	Dosearilog: Dosearilog	-2.073	0.042
High dose vs. control	Dosetreat: 1	-4.778	0.000

Just to report a significant trend test is one story, a different one is to demonstrate a significant trend and a high dose vs. control test (11).

6

3.6 Multiple endpoints

Although multiple endpoints occur frequently in long-term bioassays, such as hematology, commonly they are analysed univariately and independently. A simultaneous test over correlated endpoints can be formulated, to avoid a too high false positive decision rate (8). As an example, the changes in the activities of GSH-R levels in the testicular tissues of male rabbits after treatment with a daily dose of four nitrosamines (MEN, DEN, DMN, DPN) (17) are used. In their Table 1, the bivariate data set of GSH content and GSH-R activity was selected, assuming a correlation of 0.9 (no such correlation information available).

Table 9: Bivariate Dunnett Test

Comparisons	Test statistic	p-value
HL: den - con	-23.3114	0.0000
HL: dmn - con	-18.3434	0.0000
HL: dpn - con	-21.9466	0.0000
HL: men - con	-31.0091	0.0000
HR: den - con	-28.5824	0.0000
HR: dmn - con	-17.8634	0.0000
HR: dpn - con	-27.0118	0.0000
HR: men - con	-16.2251	0.0000

Seriously decreasing levels for both GSH content and GSH-R activity in each of the nitrosamine compounds against control exists under the assumption of a correlated bivariate normal distribution.

3.7 A counterexample where normal distribution assumption is unrealistic

Many endpoints in toxicology are not continuous, they are discrete such as counts, e.g., number of micronuclei or proportions, such as tumor rates. For these discrete endpoints the limitation of the above approach based on the normal distribution assumption is obvious. As an example the increase of mutant frequency (MF) after in utero exposure to three doses of benzopyrene in bone marrow (12) (their Figure 3) is considered where the appendix contains the summary statistics. Clearly, MF is a count and was consequently analysed in the generalized linear model in that paper assuming overdispersed Poisson counts. Although counts can be generated as well by the above approach, the limitation of normal distribution assumption becomes obvious in Figure 6 A with jittered box plots, whereas for the 20 and 40 mg group the raw data can be estimated from this figure. This makes a difference in re-analysis, see the boxplots and the evaluation for either Tukey-type trend test or Dunnett-type comparisons against control. The p-values for simulated data from mean and SD are unrealistic tiny for any comparisons, whereas ones based on estimated data are more realistic. This highlights the limitation of the ubiquitous use of arithmetic mean and SD as summary statistics – irrespective of the underlying distribution and shows the advantage of raw data presentation within boxplots (13). Comparing the the odds ratios to control (and their simultaneous confidence intervals) for both data sets, reveals a false positive tendency in the simulated data set (see appendix).
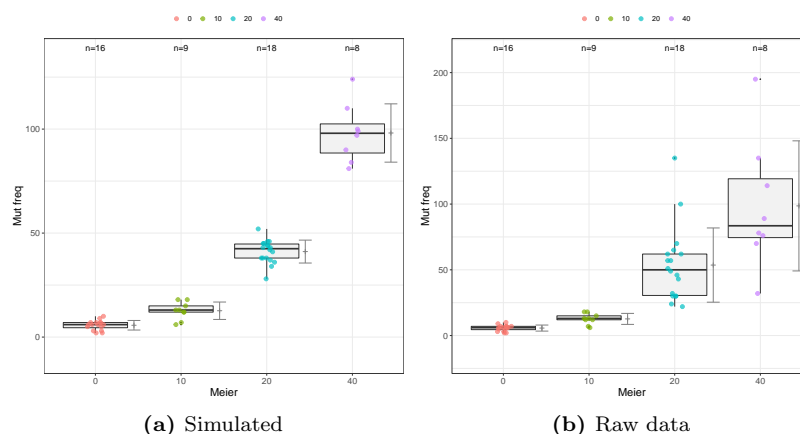

(a) Simulated   (b) Raw data

Figure 8: Box bone marrow.

4 Discussion

Based on published summary data (means, standard deviations, sample sizes) simulation-based pseudo-data can be used for statistical re-analysis. Examples are: i) Dunnett-type testing instead of independent pairwise t-tests, e.g. for NOAEL identification; ii) test statistics allowing variance heterogeneity, particularly in unbalanced designs; iii)

combining multiple bioassay studies, e.g. by using mixed effect models; iv) Analysing trends by alternative models, e.g. using dose as quantitative covariable or a combination of quanti- and qualitative formulation; v) multiple endpoint analysis (instead of independent univariate analysis) assuming low to high correlation between the endpoints; vi) using historical control data in addition to the concurrent control; vii) using robustified trend tests for possible downturn effects at high doses; viii) providing confidence intervals for the effect sites $\mu_i - \mu_0$ or μ_i/μ_0; ix) using the generated $x_{0,i}$ as part of a historical control data base; x) simulating their influence of virtual larger or smaller n_i on false negative decision rate; xi) deriving the impact of considering less dose groups, e.g. only $0, 10, 100$ instead of $0, 10, 50, 100$. Moreover, this approach can be used for teaching purposes, to establish new statistical approaches in toxicology and generating compound/strain/endpoint/sex/dose-specific data base from historical publications on individual data. For the case studies here, the data and R-code is available in the Appendix.

References

[1] V. Amrhein, S. Greenland, B. McShane, Retire statistical significance, Nature 567 (7748) (2019) 305–307 (Mar. 2019).

[2] [Anonymous]. Kick the bar chart habit. *Nature Methods*, 11(2):113–113, February 2014.

[3] J. Bunay, E. Larriba, D. Patino-Garcia, L. Cruz-Fernandes, S. Castaneda-Zegarra, M. Rodriguez-Fernandez, J. del Mazo, and R. D. Moreno. Editor's highlight: Differential effects of exposure to single versus a mixture of endocrine-disrupting chemicals on steroidogenesis pathway in mouse testes. *Toxicological Sciences*, 161(1):76–86, January 2018.

[4] N. F. Cariello and W. W. Piegorsch. The Ames test: The two-fold rule revisited. *Mutation Research-Genetic Toxicology*, 369(1-2):23–31, July 1996.

[5] Zhu Y et al. A 90-day feeding study of glyphosate-tolerant maize with the g2-aroa gene in sprague-dawley rats. *Food Chem Tox*, 2013.

[6] A. J. Fosang and R. J. Colbran. Transparency is the key to quality. *Journal of Biological Chemistry*, 290(50):29692–29694, December 2015.

[7] M. Hasler and L. A. Hothorn. Multiple contrast tests in the presence of heteroscedasticity. *Biometrical Journal*, 50(5):793–800, October 2008.

[8] M. Hasler and L. A. Hothorn. Multi-arm trials with multiple primary endpoints and missing values. *Statistics in Medicine*, 37(5):710–721, February 2018.

[9] L. A. Hothorn and G. D. Djira. A ratio-to-control williams-type test for trend. *Pharmaceutical Statistics*, 10(4):289–292, July 2011.

[10] A. Kitsche and L. A. Hothorn. Testing for qualitative interaction using ratios of treatment differences. *Statistics in Medicine*, 33(9):1477–1489, April 2014.

[11] K. K. Lin and M. A. Rahman. Comparisons of false negative rates from a trend test alone and from a trend test jointly with a control-high groups pairwise test in the determination of the carcinogenicity of new drugs. *J. Biopharm Statist*, 2018.

[12] M. J. Meier, J. M. O'Brien, M. A. Beal, B. Allan, C. L. Yauk, and F. Marchetti. In utero exposure to benzo[a]pyrene increases mutation burden in the soma and sperm of adult mice. *Environmental Health Perspectives*, 125(1):82–88, January 2017.

[13] P. Pallmann and L. A. Hothorn. Boxplots for grouped and clustered data in toxicology. *Archives of Toxicology*, 90(7):1631–1638, July 2016.

[14] B.D. Ripley. *Stochastic simulation*. Wiley, 1987.

[15] Schaarschmidt. Multiplicity adjustment for tukeys trend test and extensions to the multiple marginal generalized linear and linear mixed models.

[16] F. Schaarschmidt. Simultaneous confidence intervals for multiple comparisons among expected values of log-normal variables. *Computational Statistics and Data Analysis*, 58:265–275, FEB 2013.

[17] S. A. Sheweita, Y. Y. El Banna, M. Balbaa, I. A. Abdullah, and H. E. Hassan. N-nitrosamines induced infertility and hepatotoxicity in male rabbits. *Environmental Toxicology*, 32(9):2212–2220, September 2017.

[18] K. A. Shipkowski, C. M. Sheth, M. J. Smith, M. J. Hooth, K. L. White, and D. R. Germolec. Assessment of immunotoxicity in female fischer 344/n and sprague dawley rats and female b6c3f1 mice exposed to hexavalent chromium via the drinking water. *Journal of Immunotoxicology*, 14(1):215–227, November 2017.

[19] Z. L. Sun and S. J. et al.. Li. Alterations in epididymal proteomics and antioxidant activity of mice exposed to fluoride. *Archives of Toxicology*, 92(1):169–180, January 2018.

[20] A. Swain, J. Turton, C. Scudamore, D. Maguire, I. Pereira, S. Freitas, R. Smyth, M. Munday, C. Stamp, M. Gandhi, S. Sondh, H. Ashall, I. Francis, J. Woodfine, J. Bowles, and M. York. Nephrotoxicity of hexachloro-1:3-butadiene in the male hanover wistar rat; correlation of minimal histopathological changes with biomarkers of renal injury. *Journal of Applied Toxicology*, 32(6):417–428, June 2012.

[21] T. L. Weissgerber, V. D. Garovic, S. J. Winham, N. M. Milic, and E. M. Prager. Transparent reporting for reproducible science. *Journal of Neuroscience Research*, 94(10):859–864, October 2016.

5   Appendix: R-Code

```
## ----include=FALSE----------------------------------------------------
library(knitr)
opts_chunk$set(
cache=TRUE
)


## ----case0, echo=TRUE, result="asis", cache=TRUE, warning=FALSE----------
library(SimComp) #  laoding the selected CRAN package
set.seed(127711) # selecting a seed allows reproducibility
control <- rmvnorm(n=6, mean=6.62, sd=2.01*sqrt(6))
dose1 <- rmvnorm(n=6, mean=2.84, sd=0.99*sqrt(6))
dose2 <- rmvnorm(n=6, mean=3.30, sd=0.23*sqrt(6))
dose3 <- rmvnorm(n=6, mean=1.98, sd=0.44*sqrt(6))
dose <- factor(rep(c("0","25","50","100"), c(6,6,6,6)))
TAOC <- c(control,dose1,dose2,dose3)
datTA <- data.frame(TAOC, dose)


## ----intro, echo=FALSE, result="asis", fig.cap='Box plots of T-AOC pseudo-data.', out.width='.29\\linewidth', fig.pos="H", fig.align='center', warning=FALSE, dev='postscript', cache=TRUE----
library(toxbox)
boxclust(data=datTA, outcome="TAOC", treatment="dose", xlabel="Dose", ylabel="T-AOC", psize=1.5, hjitter=0.1, vlines="fg", white=TRUE, printN=FALSE)


## ----case17aa, echo=FALSE, result="asis", cache=TRUE, warning=FALSE-----
library(SimComp)
library(ggplot2)
set.seed(1277)
control <- rmvnorm(n=8, mean=c(50.4,50.0),  sd=c(0.1,0.2),  corr=rcm(ncol=2) )
d1      <- rmvnorm(n=8, mean=c(50.3,49.8),  sd=c(0.2,0.2),  corr=rcm(ncol=2) )
d2      <- rmvnorm(n=8, mean=c(50.0,49.4),  sd=c(0.2,0.2),  corr=rcm(ncol=2) )
d3      <- rmvnorm(n=8, mean=c(49.6,49.2),  sd=c(0.2,0.1),  corr=rcm(ncol=2) )
d4      <- rmvnorm(n=8, mean=c(49.2,49.0),  sd=c(0.1,0.1),  corr=rcm(ncol=2) )
d5      <- rmvnorm(n=8, mean=c(47.4,47.2),  sd=c(0.2,0.2),  corr=rcm(ncol=2) )
library("reshape2")
cont <- melt(control,  variable.name="Study", value.name="MCV")
colnames(cont)<- c("ID", "Study", "MCV")
cont$Dose<-c(rep("0",8))
D1 <- melt(d1,  variable.name="Study", value.name="MCV")
colnames(D1)<- c("ID", "Study", "MCV")
D1$Dose<-c(rep("15.6", 8))
D2 <- melt(d2,  variable.name="Study", value.name="MCV")
colnames(D2)<- c("ID", "Study", "MCV")
D2$Dose<-c(rep("31.3", 8))
D3 <- melt(d3,  variable.name="Study", value.name="MCV")
colnames(D3)<- c("ID", "Study", "MCV")
D3$Dose<-c(rep("62.5", 8))
D4 <- melt(d4,  variable.name="Study", value.name="MCV")
colnames(D4)<- c("ID", "Study", "MCV")
D4$Dose<-c(rep("125", 8))
D5 <- melt(d5,  variable.name="Study", value.name="MCV")
colnames(D5)<- c("ID", "Study", "MCV")
D5$Dose<-c(rep("250",8))
datmcv<-rbind(cont,D1,D2,D3,D4,D5)
datmcv$dose <- factor(datmcv$Dose, levels=c("0", "15.6", "31.3", "62.5","125","250"))


## ----case17c, echo=FALSE, result="asis", fig.cap='Box plots of MCV pseudo-data.', out.width='.29\\linewidth', fig.pos="H", fig.align='center', warning=FALSE, dev='postscript', cache=TRUE----
library(toxbox)
boxclust(data=datmcv, outcome="MCV", treatment="dose",cluster="Study",  xlabel="Dose", ylabel="MCV", psize=1.5, hjitter=0.1, vlines="fg", white=TRUE, printN=FALSE)


## ----case17f, echo=FALSE, results='asis', warning=FALSE, message=FALSE----
library(ggplot2)
library(multcomp)
library("nlme")
#mA<-lm(MCV~dose+Study+dose*Study, data=datmcv) # global IA
#anova(mA)
datmcv1<-datmcv[datmcv$Study==1, ]
datmcv2<-datmcv[datmcv$Study==2, ]
m11<-lm(MCV~dose, data=datmcv1)
m12<-lm(MCV~dose, data=datmcv2)
p11 <-summary(glht(m11,linfct= mcp(dose = "Dunnett")))
p12 <-summary(glht(m12,linfct= mcp(dose = "Dunnett")))
m1 <-lm(MCV ~ dose+Study, data=datmcv)
mm1 <- lme(MCV ~ dose+Study, random=~1|Study , data=datmcv)
dfMM <-anova(mm1)$denDF[2]
pD2 <-summary(glht(mm1,linfct= mcp(dose = "Dunnett"),df=dfMM))
P11 <- fortify(p11)[, c(1,5,6)]
P12 <- fortify(p12)[, c(5,6)]
P2 <- fortify(pD2)[, c(5,6)]
pp<-cbind(P11,P12,P2)
library(xtable)
colnames(pp)<-c("Comparisons", "$Effect size 1", "$p$-value 1", "$Effect size 2", "$p$-value 2", "$Effect size mixed", "$p$-value mixed")
print(xtable(pp, digits=c(0,1,1,5,1,5,1,5), caption="Effect sizes and $p$-values for Dunnett tests", label="tab:RE"),include.rownames=TRUE, caption.placement = "top",size="\\small",
      sanitize.text.function=function(x){x})


## ----case17h, echo=FALSE, results='asis', warning=FALSE, message=FALSE----
m1 <-lm(MCV ~ dose:Study, data=datmcv)
datmcv$study<-as.factor(datmcv$Study)
datmcv$IA<-datmcv$study:datmcv$dose
library("statint")
ialit <-iacontrast(datmcv$dose, datmcv$study, typea = "Dunnett", typeb = "Tukey")

library("nlme")
mixM <- lme(MCV ~ IA-1, random=~1|ID , data=datmcv)
plitIA <-summary(glht(mixM, linfct=ialit$cmab, df=mixM$fixDF$terms))
plit <- fortify(plitIA)[, c(1,3,6)]
library(xtable)
colnames(plit) <-c("Interaction contrast", "Effect size", "$p$-value")
print(xtable(plit, digits=c(1,1,2,3), caption="Interaction contrasts intervals", label="tab:IAD"),include.rownames=TRUE, caption.placement = "top",size="\\small",
      sanitize.text.function=function(x){x})
```

```
##␣----case3,␣echo=FALSE,␣result="asis",␣fig.cap='Box␣plot␣urine␣excretion␣pseudo␣data.',␣out.width='.29\\linewidth',␣fig.pos="H",␣fig.align='center',␣cache=TRUE,␣warning=FALSE,␣dev='postscript'----
library(SimComp)
set.seed(127711)
c␣␣<-␣ermvnorm(n=3␣␣,mean=9.3,sd=1.6)
d1␣<-␣ermvnorm(n=3␣␣,mean=9.0,sd=1.5)
d2␣<-␣ermvnorm(n=3␣␣,mean=12.7,sd=1.1)
d3<-␣ermvnorm(n=3␣␣,mean=18.2,sd=0.8)
d4␣<-␣ermvnorm(n=3␣␣,mean=22.2,sd=5.3)
d5␣<-␣ermvnorm(n=3␣␣,mean=45.8,sd=22.6)
d6␣<-␣ermvnorm(n=3␣␣,mean=30.8,sd=26.9)
d7␣<-␣ermvnorm(n=3␣␣,mean=38.1,sd=29.7)
dose␣<-␣factor(rep(c("0",␣"5",␣"10","15","20","30","45","90"),c(3,3,3,3,3,3,3,3)))
UE<-c(c,d1,d2,d3,d4,d6,d7)
datUE<-␣data.frame(UE,dose)
datUE$dose␣<-␣factor(datUE$dose,␣levels=c("0",␣"5","10","15","20","30","45",␣"90"))
library(toxbox)
boxclust(data=datUE,␣outcome="UE",␣treatment="dose",␣xlabel="",␣ylabel="urine␣excretion",␣psize=2,␣hjitter=0.01,␣vlines="fg",␣white=TRUE,␣printN=FALSE)


##␣----case3a,␣echo=FALSE,␣results="asis",␣warning=FALSE-----------------
library(mratios)
UE11<-simtest.ratioVH(UE~dose,␣data=datUE,␣type="Dunnett")
library("xtable")
ue11E␣<-as.data.frame(UE11$estimate)
ue11R␣<-as.data.frame(UE11$p.value.raw)
ue11A␣<-as.data.frame(UE11$p.value.adj)
ue11<-cbind(ue11E,ue11R,ue11A)
colnames(ue11)␣<-c("Fold␣change",␣"unadjusted␣p-value",␣"adjusted␣p-value")
xtable(ue11,␣digits=c(1,2,3,3),␣caption="Ratio-to-control␣inference␣for␣$n=3$.",
␣␣␣␣␣␣␣␣label="tab:RtC",␣caption.placement␣=␣"top",␣size="\\small")


##␣----case3b,␣echo=FALSE,␣results="asis",␣warning=FALSE-----------------
library(SimComp)
options(max.print=10000)
set.seed(127711)
c␣␣<-␣ermvnorm(n=6␣␣,mean=9.3,sd=1.6)
d1␣<-␣ermvnorm(n=6,mean=9.0,sd=1.5)
d2␣<-␣ermvnorm(n=6,mean=12.7,sd=1.1)
d3<-␣ermvnorm(n=6␣␣,mean=18.2,sd=0.8)
d4␣<-␣ermvnorm(n=6␣␣,mean=22.2,sd=5.3)
d5␣<-␣ermvnorm(n=6␣␣,mean=45.8,sd=22.6)
d6␣<-␣ermvnorm(n=6␣␣,mean=30.8,sd=26.9)
d7␣<-␣ermvnorm(n=6,mean=38.1,sd=29.7)
dose␣<-␣factor(rep(c("0",␣"5","10","15","20","30","45",␣"90"),c(6,6,6,6,6,6,6,6)))
UE<-c(c,d1,d2,d3,d4,d5,d6,d7)
datUE<-␣data.frame(UE,dose)
datUE$dose␣<-␣factor(datUE$dose,␣levels=c("0",␣"5","10","15","20","30","45",␣"90"))
library(mratios)
UE11<-simtest.ratioVH(UE~dose,␣data=datUE,␣type="Dunnett")
library(xtable)
ue11E␣<-as.data.frame(UE11$estimate)
ue11R␣<-as.data.frame(UE11$p.value.raw)
ue11A␣<-as.data.frame(UE11$p.value.adj)
ue11<-cbind(ue11E,ue11R,ue11A)
colnames(ue11)␣<-c("Fold␣change",␣"unadjusted␣$p$-value",␣"adjusted␣$p$-value")
xtable(ue11,␣digits=c(1,2,3,3),␣caption="Ratio-to-control␣inference␣for␣$n=6$.",
␣␣␣␣␣␣␣␣label="tab:RtC5",␣caption.placement␣=␣"top",␣size="\\small")


##␣----case2d,␣echo=FALSE,␣result="asis",␣fig.cap='Box␣plot␣body␣weights.',␣out.width='.29\\linewidth',␣fig.pos="H",fig.align='center',␣cache=TRUE,␣warning=FALSE,dev='postscript'----
library(SimComp)
options(max.print=10000)
set.seed(127711)
Control␣␣<-␣ermvnorm(n=12␣␣,mean=22.41,sd=1.66)#*sqrt(12))
DEHP␣<-␣ermvnorm(n=10␣␣,mean=24.93,sd=1.14)#*sqrt(10))
DBP␣<-␣ermvnorm(n=10␣␣,mean=25.19,sd=0.45)#*sqrt(10))
BBP<-␣ermvnorm(n=10␣␣,mean=25.36,sd=1.08)#*sqrt(10))
NP␣<-␣ermvnorm(n=11␣␣,mean=21.90,sd=3.55)#*sqrt(11))
OP␣<-␣ermvnorm(n=7␣␣,mean=26.18,sd=1.84)#*sqrt(7))
Mixture␣<-␣ermvnorm(n=6␣␣,mean=25.27,sd=2.57)#*sqrt(6))
dose␣<-␣factor(rep(c("Control",␣"DEHP",␣"DBP",␣"BBP",␣"NP","OP",␣"Mixture"),c(12,10,10,10,11,7,6)))
BW<-c(Control,DEHP,DBP,BBP,NP,OP,Mixture)
datBW␣<-␣data.frame(BW,dose)
datBW$dose␣<-␣factor(datBW$dose,␣levels=c("Control",␣"DEHP",␣"DBP",␣"BBP",␣"NP","OP",␣"Mixture"))
library(toxbox)
boxclust(data=datBW,␣outcome="BW",␣treatment="dose",␣xlabel="ArchTox2017",␣ylabel="TAOC",␣psize=2,␣hjitter=0.01,␣vlines="fg",␣white=TRUE,␣printN=FALSE)


##␣----case2a,␣echo=FALSE,␣results="asis",␣warning=FALSE-----------------
library(sandwich)
library(multcomp)
mBW<-lm(BW~dose,␣data=datBW)
bh<-summary(glht(mBW,␣linfct␣=␣mcp(dose␣=␣"Dunnett")))
bv<-summary(glht(mBW,␣linfct␣=␣mcp(dose␣=␣"Dunnett"),vcov␣=␣vcovHC))
bvh␣<-␣fortify(bh)[,␣c(1,6)]
bvv␣<-␣fortify(bv)[,␣c(6)]
bv11<-cbind(bvh,bvv)#␣Dunnett␣simulated␣data
library(xtable)
#bw11<-simtest.ratioVH(BW~dose,␣data=datBW,␣type="Dunnett")$p.value.adj
colnames(bv11)␣<-c("Comparison",␣"Dunnett",␣"Dunnett␣var␣het")
print(xtable(bv11,␣digits=3,␣caption="Modified␣Dunnett␣test␣allowing␣heterogeneous␣variances",␣label="tab:case2"),include.rownames=TRUE,␣caption.placement␣=␣"top",size="\\small",
␣␣␣␣␣␣sanitize.text.function=function(x){x})


##␣----case26,␣echo=FALSE,␣result="asis",␣fig.cap='Boxplot␣CHO␣endpoint␣in␣female␣rats.',␣out.width='.29\\linewidth',␣fig.pos="H",␣fig.align='center',␣cache=TRUE,␣warning=FALSE,␣dev='postscript'----
library(SimComp)
options(max.print=10000)
set.seed(1277)
control␣<-␣ermvnorm(n=10␣␣,mean=2.03,␣sd=0.32)
dose1␣<-␣ermvnorm(n=10␣␣,mean=2.87,␣sd=0.28)
dose2␣<-␣ermvnorm(n=10␣␣,mean=2.31,␣sd=0.26)
dose3␣<-␣ermvnorm(n=10␣␣,mean=2.5,␣sd=0.32)
dose␣<-␣factor(rep(c("0","12.5","25","50"),c(10,10,10,10)))
CHO<-c(control,dose1,dose2,dose3)
datCHO␣<-␣data.frame(CHO,dose)
library(toxbox)
boxclust(data=datCHO,␣outcome="CHO",␣treatment="dose",␣xlabel="Dose",␣ylabel="CHO",␣psize=2,␣hjitter=0.1,␣vlines="fg",␣white=TRUE,␣printN=FALSE)


##␣----Exa26b,␣echo=TRUE,␣results='hide',␣warning=FALSE,␣message=FALSE------
library(tukeytrend)
datCHO$Dose␣<-␣as.numeric(as.character(datCHO$dose))
fitCHO<-lm(CHO~Dose,␣data=datCHO)
ttcho<-␣tukeytrendfit(fitCHO,␣dose="Dose",␣ddf="residual",
␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣scaling=c("ari",␣"ord",␣"log",␣"arilog",␣"treat"),
␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣ctype="Williams")
exacho<-summary(glht(ttcho$mmm,␣ttcho$mlf,␣alternative="greater"))


##␣----Exam8a,␣echo=FALSE,␣results='asis',␣message=FALSE,␣warning=FALSE----
```

10

```
library(xtable)
ExaCHO <- fortify(exacho)[,c(1,5,6)]
colnames(ExaCHO) <- c("Dose metameter", "Test statistic", "$p$-value")
print(xtable(ExaCHO, digits=4, caption="Tukey and Williams Trend Test",
       label="tab:exacho"),include.rownames=FALSE, caption.placement = "top",size="\\small",
       sanitize.text.function=function(x){x})


## ----case6, echo=FALSE, result="asis", fig.cap='Boxplot VAP.', out.width='.29\\linewidth', fig.pos="H", cache=TRUE,fig.align='center', warning=FALSE, dev='postscript'----
library(SimComp)
options(max.print=10000)
set.seed(1277)
control <- ermvnorm(n=3  ,mean=94.37,sd=1.85*sqrt(3))
dose1 <- ermvnorm(n=3  ,mean=84.16,sd=2.54*sqrt(3))
dose2 <- ermvnorm(n=3  ,mean=89.99,sd=1.45*sqrt(3))
dose3 <- ermvnorm(n=3  ,mean=76.55,sd=3.98*sqrt(3))
dose <- factor(rep(c("0", "0.05", "5", "50"),c(3,3,3,3)))
Dose <- as.numeric(rep(c("0", "0.05", "5", "50"),c(3,3,3,3)))
VAP<-c(control,dose1,dose2,dose3)
datVA <- data.frame(VAP,dose, Dose)
library(toxbox)
boxclust(data=datVA, outcome="VAP", treatment="dose", xlabel="Dose", ylabel="VAP", psize=2, hjitter=0.1, vlines="fg", white=TRUE, printN=FALSE)


## ----case6a, echo=FALSE, results="asis", warning=FALSE------------------
library(tukeytrend)
cm <- matrix(c(-1,0,0,1), ncol=4)
fitVAP<-lm(VAP~Dose, data=datVA)
ttvap <- tukeytrendfit(fitVAP, dose="Dose", ddf="residual",
                       scaling=c("ari", "ord", "arilog", "treat"),  ctype=cm)
exavap<-summary(glht(ttvap$mmm, ttvap$mlf,alternative="less"))

library(xtable)
ExaVAP <- fortify(exavap)[,c(1,5,6)]
rownames(ExaVAP) <- c("Trend with arithmetic scores", "Trend with ordinal scores", "Trend with logarithmic scores", "High dose vs. control")
colnames(ExaVAP) <- c("Comparison", "Test statistic", "$p$-value")
print(xtable(ExaVAP, digits=3, caption="User-defined comparisons", label="tab:case6a"),include.rownames=TRUE, caption.placement = "top",size="\\small",
       sanitize.text.function=function(x){x})


## ----case30, echo=FALSE, results="asis", warning=FALSE------------------
library(SimComp)
set.seed(5678)
con <- ermvnorm(n=5,mean=c(7.35,106.5),sd=c(0.38,4.68),corr=rbind(c(1,0.9),c(0.9,1)))
men <- ermvnorm(n=5,mean=c(1.67,70.55),sd=c(0.14,4.35),corr=rbind(c(1,0.9),c(0.9,1)))
den <- ermvnorm(n=5,mean=c(3.08,43.17),sd=c(0.23,2.36),corr=rbind(c(1,0.9),c(0.9,1)))
dmn <- ermvnorm(n=5,mean=c(3.99,66.92),sd=c(0.27,2.19),corr=rbind(c(1,0.9),c(0.9,1)))
dpn <- ermvnorm(n=5,mean=c(3.33,46.65),sd=c(0.36,3.19),corr=rbind(c(1,0.9),c(0.9,1)))
treat <- factor(rep(c("con", "men", "den", "dmn", "dpn"),c(5,5,5,5,5)))
GSHL<-c(con[,1],men[,1],den[,1],dmn[,1],dpn[,1])
GSHR<-c(con[,2],men[,2],den[,2],dmn[,2],dpn[,2])
datSHE <- data.frame(GSHL, GSHR,treat)


## ----case30a, echo=FALSE, results="asis", warning=FALSE-----------------
library(multcomp)
library(sandwich)
mL<-lm(GSHL~treat, data=datSHE)
mR<-lm(GSHR~treat, data=datSHE)
exLR <- summary(glht(mmm(HL = mL, HR = mR), mlf(mcp(treat ="Dunnett"))))


## ----Exam30c, echo=FALSE, results='asis', message=FALSE, warning=FALSE----
library(ggplot2)
library(xtable)
exaLR <- fortify(exLR)[,c(1,5,6)]
colnames(exaLR) <- c("Comparisons", "Test statistic", "$p$-value")
print(xtable(exaLR, digits=4, caption="Bivariate Dunnett Test",
       label="tab:ex30c"),include.rownames=FALSE, caption.placement = "top",size="\\small",
       sanitize.text.function=function(x){x})


## ----case5, echo=FALSE, result="asis", fig.cap='Box bone marrow.',fig.subcap=c('Simulated', 'Raw data'), out.width='.29\\linewidth',fig.align='center',fig.pos="H", cache=TRUE, warning=FALSE, dev='postscript'
library(SimComp)
options(max.print=10000)
set.seed(1277)
control <- ermvnorm(n=16  ,mean=6.2,sd=2.19)
dose1 <- ermvnorm(n=9  ,mean=13.1,sd=4.03)
dose2 <- ermvnorm(n=18  ,mean=41.6,sd=5.46)
dose3 <- ermvnorm(n=8  ,mean=98.5,sd=14.05)
Dose2<-c(135,100,70,65,62,62,57,57,51,49,46,43,32,30,30,30,24,22)
Dose3<-c(32,70,76,78,89,114,135,195)
dose <- factor(rep(c("0", "10", "20", "40"),c(16,9,18,8)))
MF<-c(control,dose1,dose2,dose3)
mf<-c(control,dose1,Dose2,Dose3)
datMF <- data.frame(MF,dose)
datmf <- data.frame(mf,dose)
datMF$MUF<-floor(datMF$MF) # generating integer count variable
datmf$muf<-floor(datmf$mf) # generating integer count variable
library(toxbox)
boxclust(data=datMF, outcome="MUF", treatment="dose", xlabel="", ylabel="Mut freq", psize=2, hjitter=0.1, vlines="fg", white=TRUE, printN=FALSE)
boxclust(data=datmf, outcome="muf", treatment="dose", xlabel="", ylabel="Mut freq", psize=2, hjitter=0.1, vlines="fg", white=TRUE, printN=FALSE)


## ----case5b, echo=TRUE,results='asis', warning=FALSE, message=FALSE------
library(tukeytrend)
datMF$Dose<-as.numeric(as.character(datMF$dose))
datmf$Dose<-as.numeric(as.character(datmf$dose))
fitQPMF <- glm(MUF ~ Dose, data=datMF, family=quasipoisson(link="log"))
fitQPmf <- glm(muf ~ Dose, data=datmf, family=quasipoisson(link="log"))
Exa5 <- tukeytrendfit(fitQPMF, dose="Dose", scaling=c("ari", "ord", "arilog", "treat"),
                      ctype="Dunnett")
EXA5<-summary(glht(model=Exa5$mmm, linfct=Exa5$mlf))
Exa6 <- tukeytrendfit(fitQPmf, dose="Dose", scaling=c("ari", "ord", "arilog", "treat"),
                      ctype="Dunnett")
EXA6<-summary(glht(model=Exa6$mmm, linfct=Exa6$mlf))
```