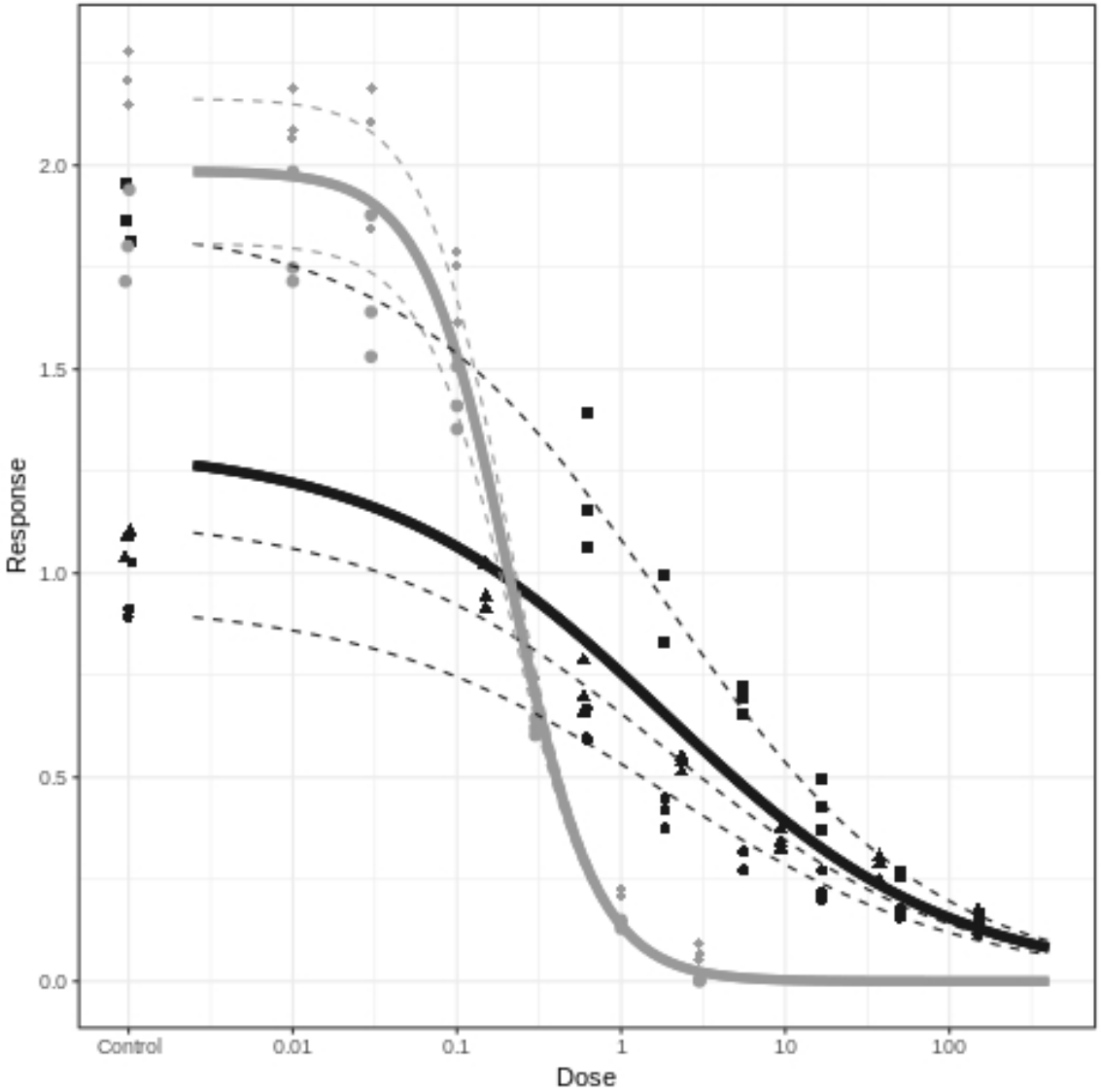
Advances in dose-response analysis

Christian Ritz & Jens C. Streibig

useR! Toulouse, July 12 2019

It's all about such data



Unifying feature: statistical models fully parametric and nonlinear (s-shape)

Short history

- Dose-response analysis already used in the 1940s and 1950s in toxicology (small experiments)
- Probit regression models fitted by means of linearization, an approximation, e.g., Finney (1971)
- In the 1980s and 1990s several stand-alone programmes and Excel/SAS macros for fitting nonlinear regression models appeared
- In 2005 the first version of the R package *drc* for fitting dose-response curves appeared (Ritz & Streibig, 2005)
- From 2005 to 2019 *drc* was modified and extended substantially, becoming an unparalled general infrastructure for dose-response analysis (Ritz *et al.*, 2019)

Current ecosystem *drc*

Meanwhile the development moved to GitHub:

https://github.com/DoseResponse

Currently, there are 4 packages:

- *drc* (still on CRAN, but not updated so often)
- *drcData* containing only data
- *medrc* for mixed dose-response models
- *bmd* for benchmark dose estimation

(another indepedent but related package on GitHub: *drcSeedGerm*)

Modular structure

A general trend in the R package landscape is creating packages that serve as modules and may be combined to provide more advanced solutions than possible using just one package

- Use specialized packages for specialized tasks

Functionality of *drc* extended through the following packages:

- *ggplot2*
- *metafor*
- *multcomp*
- *nlme*
- *sandwich*

Example 1: Binary dose-response data

Fitting a concentration-response model to dose-response data on toxicity of 4 different types of selenium (multiple curves)

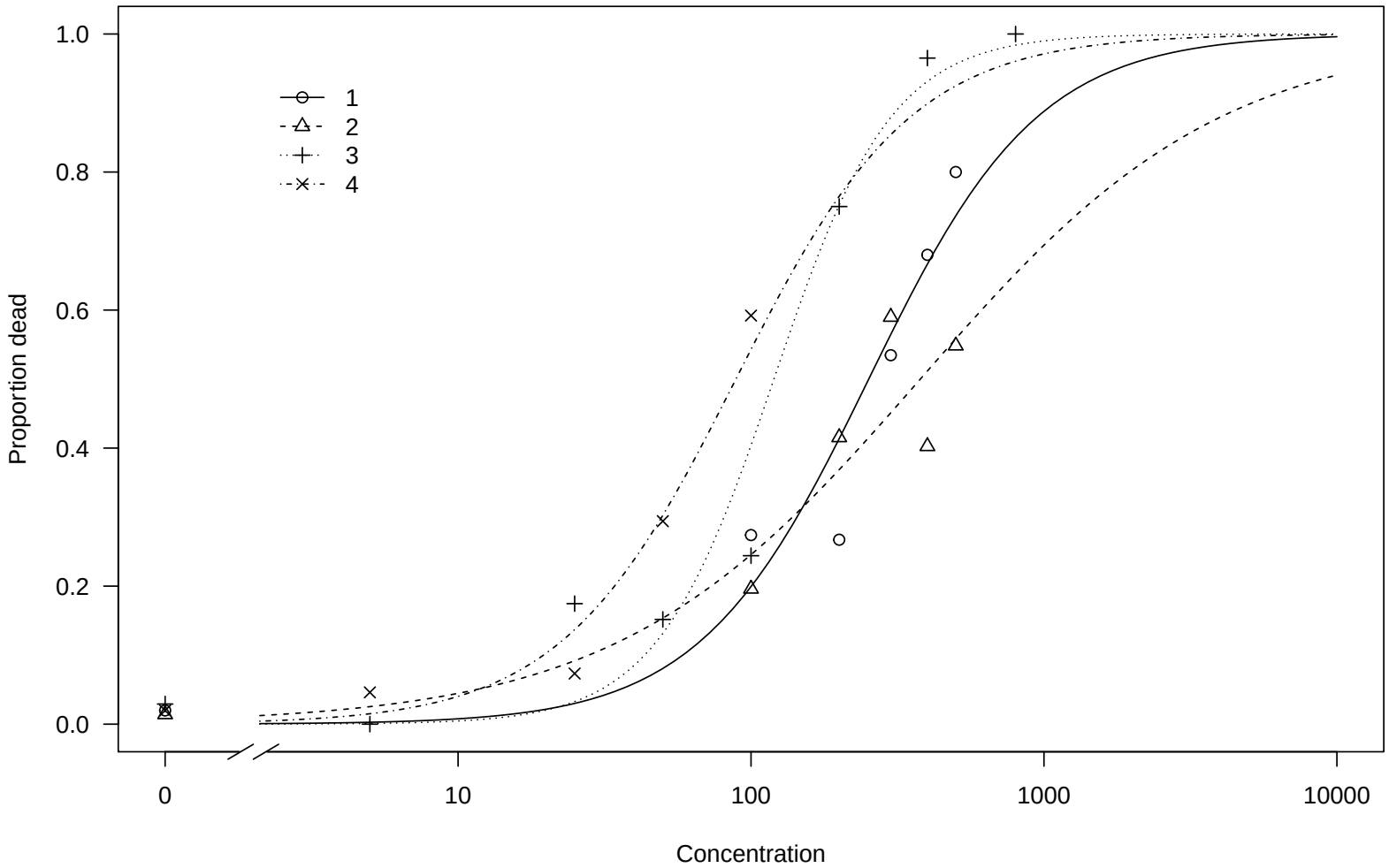
```
library(devtools)
install_github("DoseResponse/drc")
library(drc)

selenium.LL.2.1 <- drm(dead/total ~ conc,
                       curveid = type,
                       weights = total,
                       data = selenium,
                       fct = LL.2(),
                       type = "binomial")
```

Specifically, a two-parameter log-logistic model is fitted with different parameters for different types of selenium; somewhat similar specification as for glm()

The analysis of these data comprised an entire statistics article, 10 years ago (Jeske *et al.*, 2009)

Example 1: Visualization of the model fit

```
plot(selenium.LL.2.1, type = "all", broken = TRUE, xlim = c(0, 10000),
     xlab = "Concentration", ylab = "Proportion dead", legendPos = c(5, 0.95))
```



Example 1: Effective concentrations

Estimating effective concentrations with the usual (naive) standard errors and confidence intervals derived using the delta method

```
ED(selenium.LL.2.1, c(50), interval = "delta")
```

```
##
## Estimated effective doses
##
##         Estimate Std. Error    Lower    Upper
## e:1:50 252.2556    13.8268 225.1555 279.3556
## e:2:50 378.4605    39.3707 301.2953 455.6256
## e:3:50 119.7132     5.9054 108.1389 131.2875
## e:4:50  88.8053     8.6161  71.9180 105.6926
```

Example 1: Robust standard errors

Estimating effective concentrations with robust standard errors (using *sandwich*) and the delta method: the modern version of the classical adjustment for over-dispersion (multiplication by a scaling factor)

```
library(sandwich)

ED(selenium.LL.2.1, c(50), interval = "delta", vcov. = sandwich)
```

```
##
## Estimated effective doses
##
##        Estimate Std. Error   Lower   Upper
## e:1:50  252.256     27.842 197.686 306.825
## e:2:50  378.460     82.470 216.822 540.099
## e:3:50  119.713     15.457  89.419 150.008
## e:4:50   88.805     10.543  68.142 109.469
```

In this case standard errors mostly increase substantially as compared to the ones on the previous slide; appreciable model misspecification

Example 1: Simultaneous confidence intervals

Exploiting that the output may be used with *multcomp*

```
selenium.EDres <- ED(selenium.LL.2.1, c(50),
                     interval = "delta", vcov. = sandwich,
                     multcomp = TRUE, display = FALSE)

library(multcomp)
confint(glht(selenium.EDres[["EDmultcomp"]]))
```

```
##
##   Simultaneous Confidence Intervals
##
## Fit: NULL
##
## Quantile = 2.4908
## 95% family-wise confidence level
##
##
## Linear Hypotheses:
##              Estimate lwr      upr
## e:1:50 == 0 252.2556 182.9052 321.6059
## e:2:50 == 0 378.4605 173.0405 583.8804
## e:3:50 == 0 119.7132  81.2130 158.2134
## e:4:50 == 0  88.8053  62.5451 115.0655
```

Example 1: Pairwise comparisons

Even pairwise comparisons can be obtained using *multcomp*:

```
summary(glht(selenium.EDres[["EDmultcomp"]], linfct = contrMat(1:4, "Tukey")))
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Linear Hypotheses:
##             Estimate Std. Error z value Pr(>|z|)
## 2 - 1 == 0   126.20      87.04   1.450  0.43135
## 3 - 1 == 0  -132.54      31.84  -4.162  < 0.001 ***
## 4 - 1 == 0  -163.45      29.77  -5.490  < 0.001 ***
## 3 - 2 == 0  -258.75      83.91  -3.084  0.00884 **
## 4 - 2 == 0  -289.66      83.14  -3.484  0.00227 **
## 4 - 3 == 0   -30.91      18.71  -1.652  0.31469
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

This works for arbitrary EC values

Example 2: The idea first

Two steps:

1. Fitting a dose-response model to each sub-experiment and extracting the relevant parameter estimate and corresponding standard error from each model fit
2. Combining parameter estimates in a meta analysis/regression (may be univariate or multivariate)

Similar to a meta analysis/regression, but only involving a single (but complex) experiment or study

This meta-analytic approach was suggested as an alternative to dose-response (nonlinear) mixed-effects models by Jiang & Kopp-Schneider (2014) and extended to an event-time setting by Jensen *et al.* (2017)

It offers a powerful extension of the functionality of *drc*

Example 2: Step 1 – fitting separate models

```
install_github("DoseResponse/drcData")
library(drcData)
blackgrass[["Pot"]] <-
  with(blackgrass, as.numeric(interaction(Exp, Bio, Depth, Temp, Rep)))

library(plyr)
fitFct.LL.3 <- function(dataSet)
{
    modelFit <- try(drm(Ger ~ Start.Day + End.Day,
                        data = dataSet,
                        fct = LL.3(),
                        type = "event"), silent = TRUE)
    if (inherits(modelFit, "try-error")) {modelFit <- NULL}
    return(modelFit)
}

black.grass.modelfits2 <-
  dlply(blackgrass, .(Exp, Bio, Depth, Temp, Pot), fitFct.LL.3)
```

Example 2: Step 1 – fomatting results

Step 1: formatting results

```
paramFct.LL.3 <- function(fitObj)
{
  if (is.null(fitObj)) {return(rep(NA, 6))}
  # handling replicates with all values missing

  coefSum <- coef(summary(fitObj))
  returnVec <- c(coefSum[1, 1:2],  # slope
                 coefSum[2, 1:2],  # maximum
                 ED(fitObj, 50, display = FALSE)[1:2])  # t50

  names(returnVec) <- c("b", "b.se",
                        "d", "d.se",
                        "t50", "t50.se")
  returnVec
}

blackgrass.parms <- ldply(black.grass.modelfits2, paramFct.LL.3)

blackgrass.parms[["BioDepthTemp"]] <-
  with(blackgrass.parms, interaction(Bio, Depth, Temp))
```

Example 2: Step 2 – meta analysis

```
head(blackgrass.parms, 2)
```

```
##   Exp Bio Depth Temp Pot          b     b.se          d       d.se      t50
## 1   1   R     0   10   1 -7.454884 1.544786 0.6983622 0.08645511 686.5420
## 2   1   R     0   10  33 -9.852617 1.818852 0.8253547 0.06951696 708.4872
##      t50.se BioDepthTemp
## 1 36.02073       R.0.10
## 2 24.27229       R.0.10
```

```
library(metafor)
blackgrass.t50.mm <- rma.mv(t50, (t50.se)^2,
                            mods = ~ BioDepthTemp - 1,
                            random = ~ 1|Exp/Pot,
                            data = blackgrass.parms)
```

Output: Estimated means per combination; next step: pairwise comparisons using *multcomp*

Example 3: Species sensitivity distributions

This is a recent addition to *drc*, showing that yet another type of statistical analysis, estimation of species sensitivity distributions, is also a dose-response analysis

In this case a single distribution is fitted to the observed data; still the result is visualized by means of a dose-response curve

Data are 48- to 96-hour acute toxicity values (LC50 and EC50 values from dose-response analysis) for exposure of Australian and Non-Australian arthropod, nonarthropod invertebrate, fish, and amphibian to the pesticide endosulfan (Hose & Van den Brink, 2004)

Data retrieved like this:

```
library(fitdistrplus)  # to retrieve the data
data("endosulfan")
endosulfan.art <- subset(endosulfan, group == "Arthropods" & Australian == "no")
```

(we only look at the Non-Australian data)

Fitting an SSD

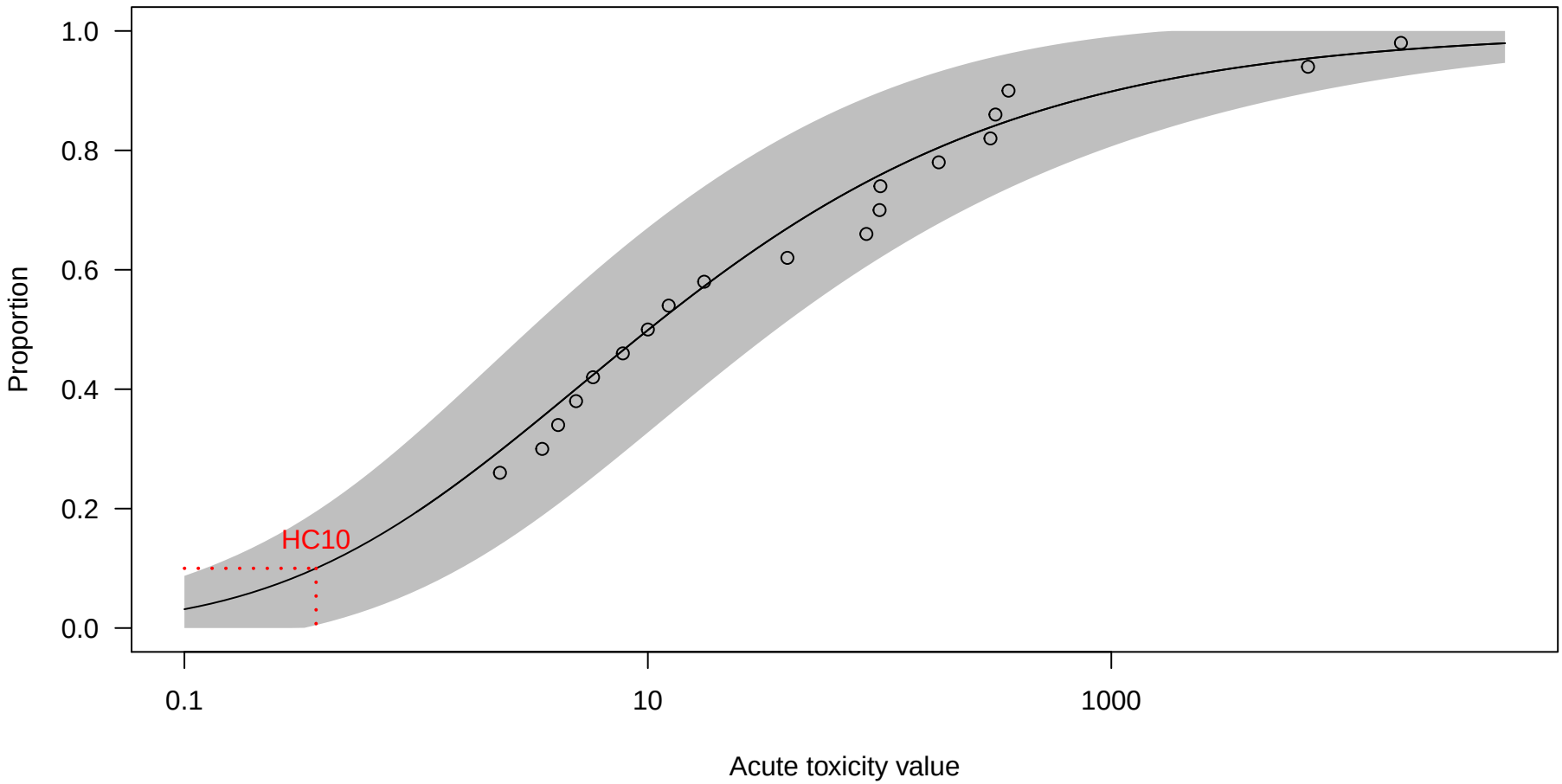
Fitting a Burr type III distribution (a generalized log-logistic distribution), which is a special case of the built-in five-parameter log-logistic model function in *drc*:

```
endo.art.no <- drm(~ ATV, data = endosulfan.art,
                   fct = LL.5(fixed = c(NA, 0, 1, NA, NA)), type = "ssd")

summary(endo.art.no)
```

```
##
## Model fitted: Generalized log-logistic (ED50 as parameter) (3 parms)
##
## Parameter estimates:
##
##                Estimate Std. Error t-value   p-value
## b:(Intercept) -0.42124    0.10130 -4.1582 3.207e-05 ***
## e:(Intercept)  0.14826    0.82215  0.1803    0.8569
## f:(Intercept)  4.43626    7.86226  0.5642    0.5726
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Example 3: Showing the fitted curve

```
plot(endo.art.no, type = "confidence", xlim = c(0.1, 50000), ylim = c(0, 1),
     xlab = "Acute toxicity value", ylab = "Proportion")
plot(endo.art.no, xlim = c(1, 50000), ylim = c(0, 1), add = TRUE)
```



Example 3: Estimated HC values

Hazard concentrations (HC) and confidence intervals estimated through inverse regression:

```
ED(endo.art.no, c(5, 10, 50))
```

```
##
## Estimated effective doses
##
##         Estimate Std. Error
## e:1:5    0.16151    0.12397
## e:1:10   0.36985    0.23723
## e:1:50  10.07614    6.52066
```

```
#ED(endo.art.no, c(5,10,50), interval = "delta")  # this you don't want to do
ED(endo.art.no, c(5, 10, 50), interval = "inv")
```

```
##
## Estimated effective doses
##
##          Estimate     Lower     Upper
## e:1:5    0.161510  0.045483  0.842860
## e:1:10   0.369848  0.122836  1.556769
## e:1:50  10.076144  3.088323 35.805902
```

Concluding remarks

Recent advances include:

- simultaneous inference
- two-step approach
- more model fitting options:
  - event-time data (left-censored)
  - species sensitivity distributions
  - mixed-effects dose-response models, e.g., Baty *et al.* (2016, 2017) and da Cunha *et al.* (2019)

Some ideas for future developments and projects:

- exploiting combined *metafor* and *sandwich* capabilities
- implementation of more general event-time models (beyond left censoring)
- biphasic models for intake of nutrients

If interested please drop me an e-mail

References

- da Cunha, B. R., Andreasen, C., Rasmussen, J., Nielsen, J., Ritz, C., Streibig, J. C. (2019). Assessing herbicide symptoms by using a logarithmic field sprayer. *Pest Management Science*, **75**, 1166–1171
- Baty, F., Ritz, C., van Gestel, A., Brutsche, M., Gerhard, D. (2016). Modeling the oxygen uptake kinetics during exercise testing of patients with chronic obstructive pulmonary diseases using nonlinear mixed models. *BMC Medical Research Methodology*, **16**, 66
- Baty, F., Ritz, C., Jensen, S. M., Kern, L., Tamm, M., Brutsche, M. (2017). Multimodel inference applied to oxygen recovery kinetics after 6-min walk tests in patients with chronic obstructive pulmonary disease. *PLOS ONE*, **12**, e0187548
- Finney, D. J. (1971). *Probit analysis*. Cambridge University Press, 3rd edition
- Hose, G. C., Van den Brink, P. (2004). Confirming the Species-Sensitivity Distribution concept for Endosulfan Using Laboratory, Mesocosm, and Field Data. *Archives of Environmental Contamination and Toxicology*, **45**, 511-520
- Jensen, S. M., Andreasen, C., Streibig, J. C., Keshtkar, E., Ritz, C. (2017). A note on the analysis of germination data from complex experimental designs. *Seed Science Research*, **27**, 321-327
- Jeske, D. R., Xu, H. K., Blessinger, T., Jensen, P., Trumble, J. (2009). Testing for the Equality of EC50 Values in the Presence of Unequal Slopes With Application to Toxicity of Selenium Types. *Journal of Agricultural, Biological, and Environmental Statistics*, **14**, 469-483
- Jiang, X., Kopp-Schneider, A. (2014) Summarizing EC50 estimates from multiple dose-response experiments: A comparison of a meta-analysis strategy to a mixed-effects model approach. *Biometrical Journal*. **56**, 493-512
- Ritz, C., Streibig, J. C. (2005). Bioassay Analysis using R. *Journal of Statistical Software*, **12**, 5. http://www.jstatsoft.org/v12/i05
- Ritz, C., Baty, F., Streibig, J. C., Gerhard, D. (2015). Dose-Response Analysis Using R. *PLoS ONE*, **10**, e0146021. https://doi.org/10.1371/journal.pone.0146021
- Ritz, C., Jensen, S. M., Gerhard, D., Streibig, J. C. (2019). Dose-Response Analysis Using R. CRC Press. *Forthcoming*. https://www.crcpress.com/Dose-Response-Analysis-Using-R/Ritz-Jensen-Gerhard-Streibig/p/book/9781138034310