Dose-response modelling - Bayesian Inference

The R-package `morse` for binary and count data

PREDITOX School

Lyon, France

Contact: sandrine.charles@univ-lyon1.fr

About `morse`

Current version: 3.3.2

Modelling Tools for Binary and Count Data in Ecotoxicology.

https://cran.r-project.org/package=morse

Baudrot, Virgile, and Sandrine Charles. 2021. `morse`: An R-Package in Support of Environmental Risk Assessment. *Journal of Open Source Software*, 6 (68): 3200. https://doi.org/10.21105/joss.03200.

Get an example data set

Survival and reproduction data from a chronic toxicity test with *Daphnia magna* exposed to six concentrations of chlordan during 21 days.

Six concentrations were tested, with 10 replicates per concentration.

Each replicate contained one organism. Survival and reproduction were monitored at 22 time points.

```
library(morse)
data(chlordan) # load raw data
```

Manar, R., Bessi, H. and Vasseur, P. (2009) Reproductive effects and bioaccumulation of chlordan in *Daphnia magna*, *Environmental Toxicology and Chemistry*, 28, 2150-2159.

Look at the raw binary data (1)

```
head(chlordan)
```

```
  conc time Nsurv Nrepro replicate
1    0    0     1      0         1
2    0    1     1      0         1
3    0    2     1      0         1
4    0    3     1      0         1
5    0    4     1      0         1
6    0    5     1      0         1
```

```
# Five columns
```

Look at the raw binary data (2)

```
dataset <- survData(chlordan)
# Create en R object to be used with 'morse'
# Check consistency of the data
# Number of replicates per time and concentration:
summary(dataset, quiet=TRUE)$NbrepTimeConc
```

```
      time
conc    0  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 1
  0    10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 1
  0.18 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 1
  0.73 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 1
  1.82 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 1
  2.9  10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 1
  7    10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 1
```

Look at the raw binary data (3)

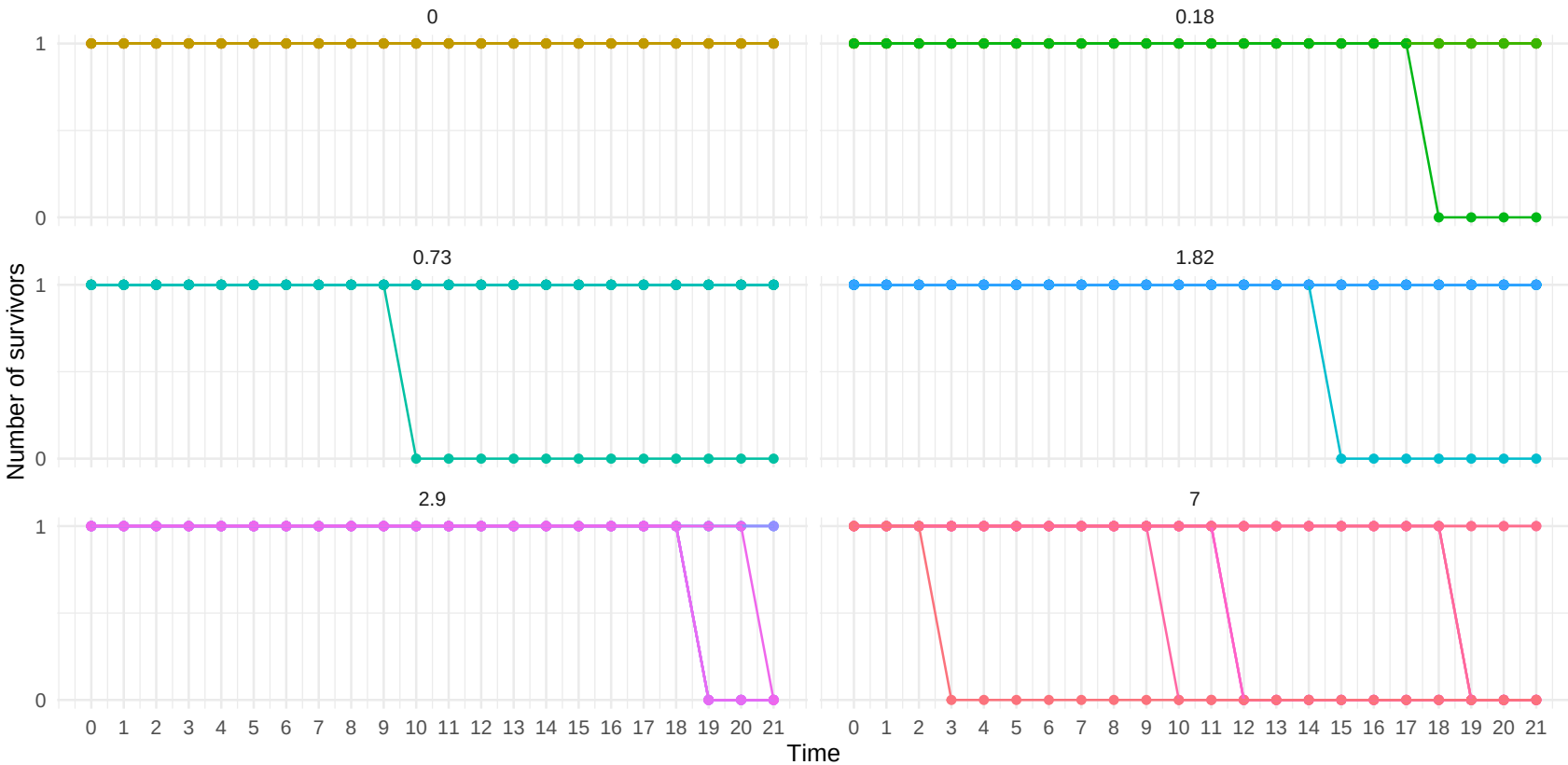
```
# Number of survivors (sum of replicates)
# per time and concentration
summary(dataset, quiet=TRUE)$NbsurvTimeConc
```

```
      0  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17
0    10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10
0.18 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10
0.73 10 10 10 10 10 10 10 10 10 10  9  9  9  9  9  9  9  9
1.82 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10  9  9  9
2.9  10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10
7    10 10 10  9  9  9  9  9  9  9  8  8  5  5  5  5  5  5
```

Plot raw binary data

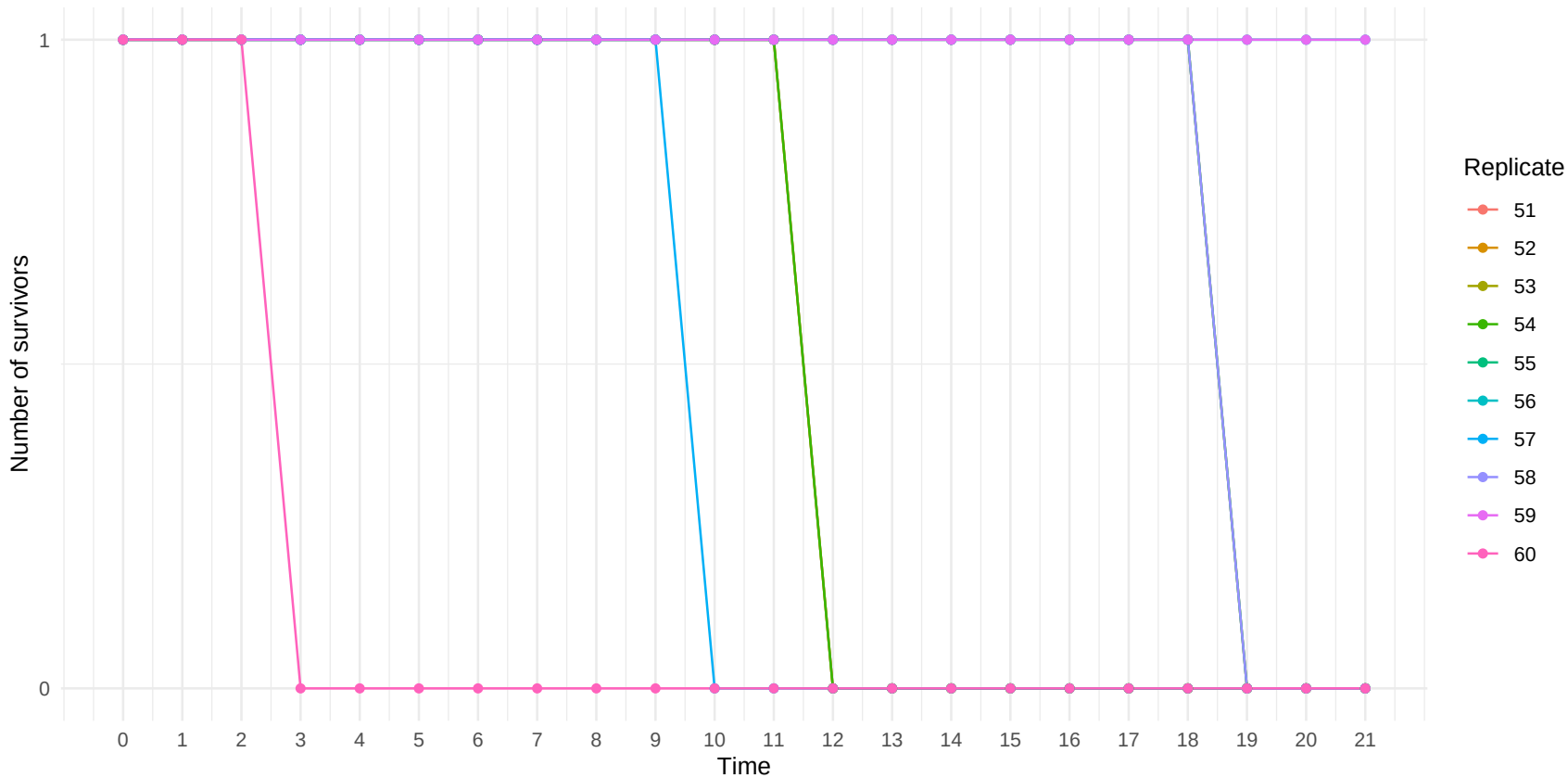
Number of surviving organisms versus time at each concentration and each replicate.

```
plot(dataset)
```



Focus on one concentration in particular

```
plot(dataset, concentration = 7, addlegend = TRUE,
     pool.replicate = FALSE)
```

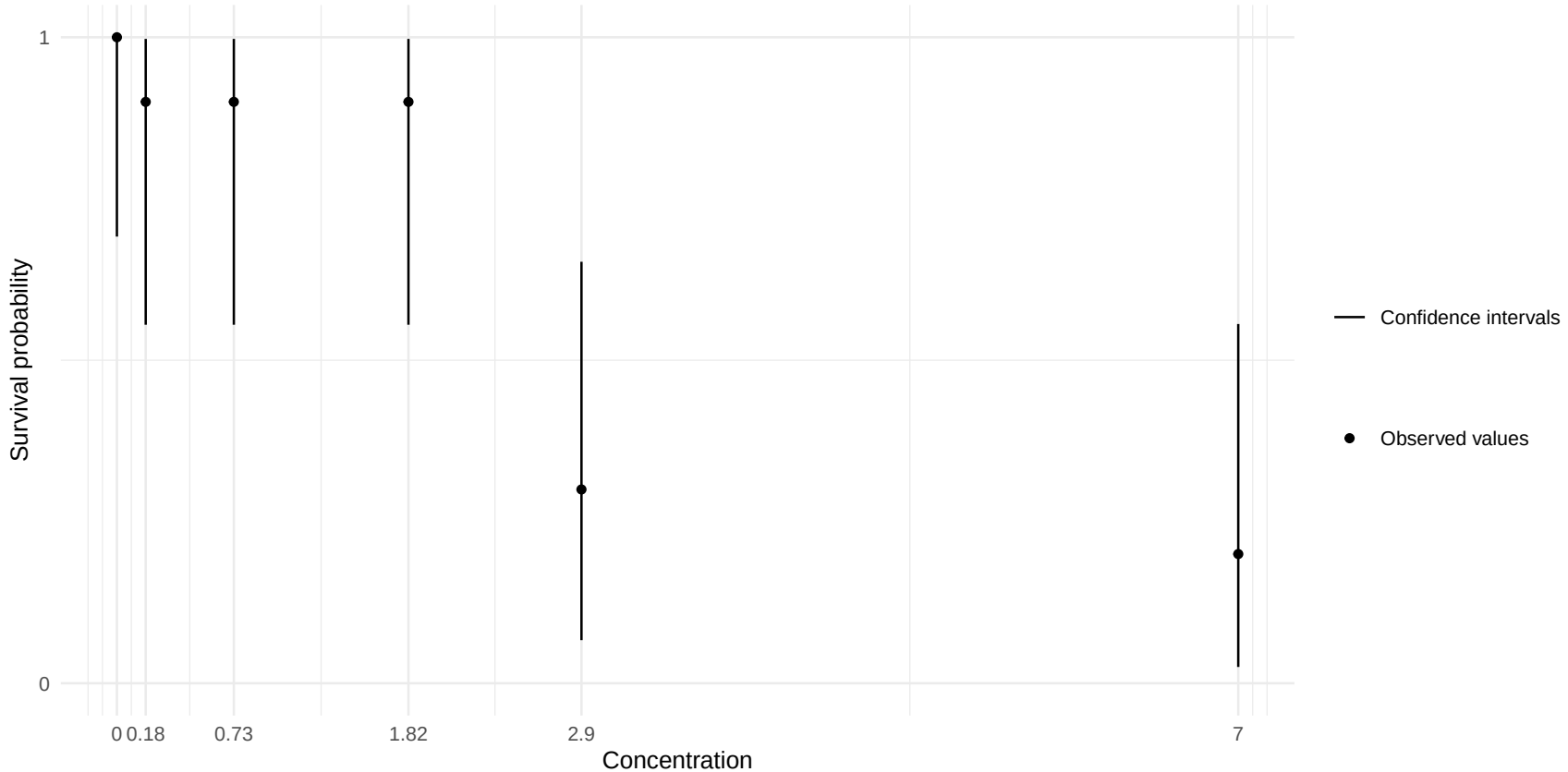


Plot the concentration-response curve at any target time

Survival rate versus concentration at day 21 (target time).

Binomial confidence intervals are added to the observed values.

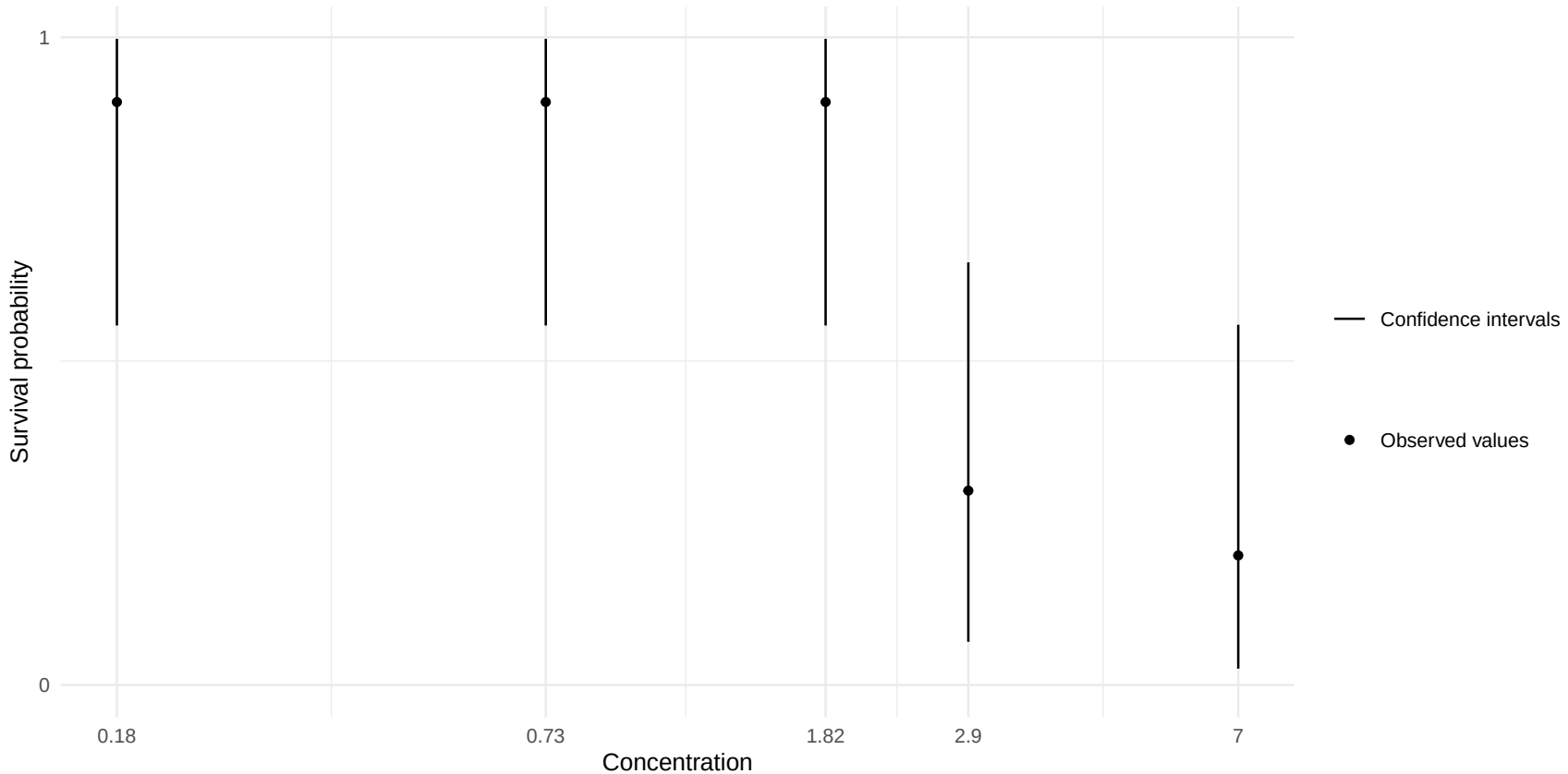
```
plotDoseResponse(dataset, target.time = 21,
                 addlegend = TRUE)
```



Plot the concentration-response curve at any target time

Change the x-scale

```
plotDoseResponse(dataset, target.time = 21,
                 addlegend = TRUE, log.scale=TRUE)
```



DR model for binary data: a reminder

In a standard analysis of binary data, a concentration-response model is fitted on data at target-time.

Binary (or quantal) data, such as for example the number N_i of surviving organisms at time t at concentration X_i, follows a binomial distribution:

$$N_i \sim \mathcal{B}(n_i^{init}, f(X_i))$$

where n_i^{init} is the initial number of organisms at concentration X_i.

In the R-package `morse`, we use a 3p-log-logistic deterministic part:

$$f(X) = \frac{d}{1 + (\frac{X}{e})^b}$$

where X stands for the contaminant concentration with $f(X)$ the probability of success (e.g., the survival probability). Parameters b, e and d are positive: d corresponds to the probability of success when $X = 0$; e corresponds to the X value for which the probability of success equals $\frac{d}{2}$ (namely, the LC_{50} for survival data); and b (the *slope* or the *curvature* of the curve) is related to the effect intensity of the contaminant.

DR model for binary data: priors

Posterior distributions for parameters b, d and e are estimated using JAGS from the following priors:

$\log_{10} e \sim \mathcal{N}(\mu_e, \sigma_e)$

with $\mu_e = \frac{\log_{10}(\min c_i)+\log_{10}(\max c_i)}{2}$

and $\sigma_e = \frac{\log_{10}(\max c_i)-\log_{10}(\min c_i)}{4}$.

$\log_{10} b \sim \mathcal{U}(-2, 2)$

If there is no failure at $X = 0$, then we fix $d = 1$, otherwise we assume $d \sim \mathcal{U}(0, 1)$.

[Forfait-Dubuc et al., 2012; Delignette-Muller et al., 2017]

DR model for binary data: fitting at day 21

```
fit <- survFitTT(dataset, quiet=TRUE,
                 target.time = 21,
                 lcx = c(10, 20, 50))
```

here the target time is fixed at 21 days:

$\rightarrow$ default value = end of experiment.

We can ask for the estimation of any LC_x values:

here $x = 10, 20, 50\%$.

DR model for binary data: parameter estimates

Get parameter estimates as medians and 95% credible intervals:

```
summary(fit, quiet=TRUE)$Qpost
```

```
         50%       2.5%      97.5%
b 1.182e+00 4.821e-01 2.121e+00
e 2.682e+00 1.528e+00 5.175e+00
```

What do you notice?

DR model for binary data: priors-posteriors

Priors

```
summary(fit, quiet=TRUE)$Qpriors[,2:3]
```

```
       2.5%     97.5%
b 1.259e-02 7.943e+01
e 1.867e-01 6.748e+00
```

Posteriors

```
summary(fit, quiet=TRUE)$Qpost[,2:3]
```

```
       2.5%     97.5%
b 4.821e-01 2.121e+00
e 1.528e+00 5.175e+00
```

DR model for binary data: LC_x estimates

Get LC_x estimates as medians and 95% credible intervals

```
summary(fit, quiet=TRUE)$QLCx
```

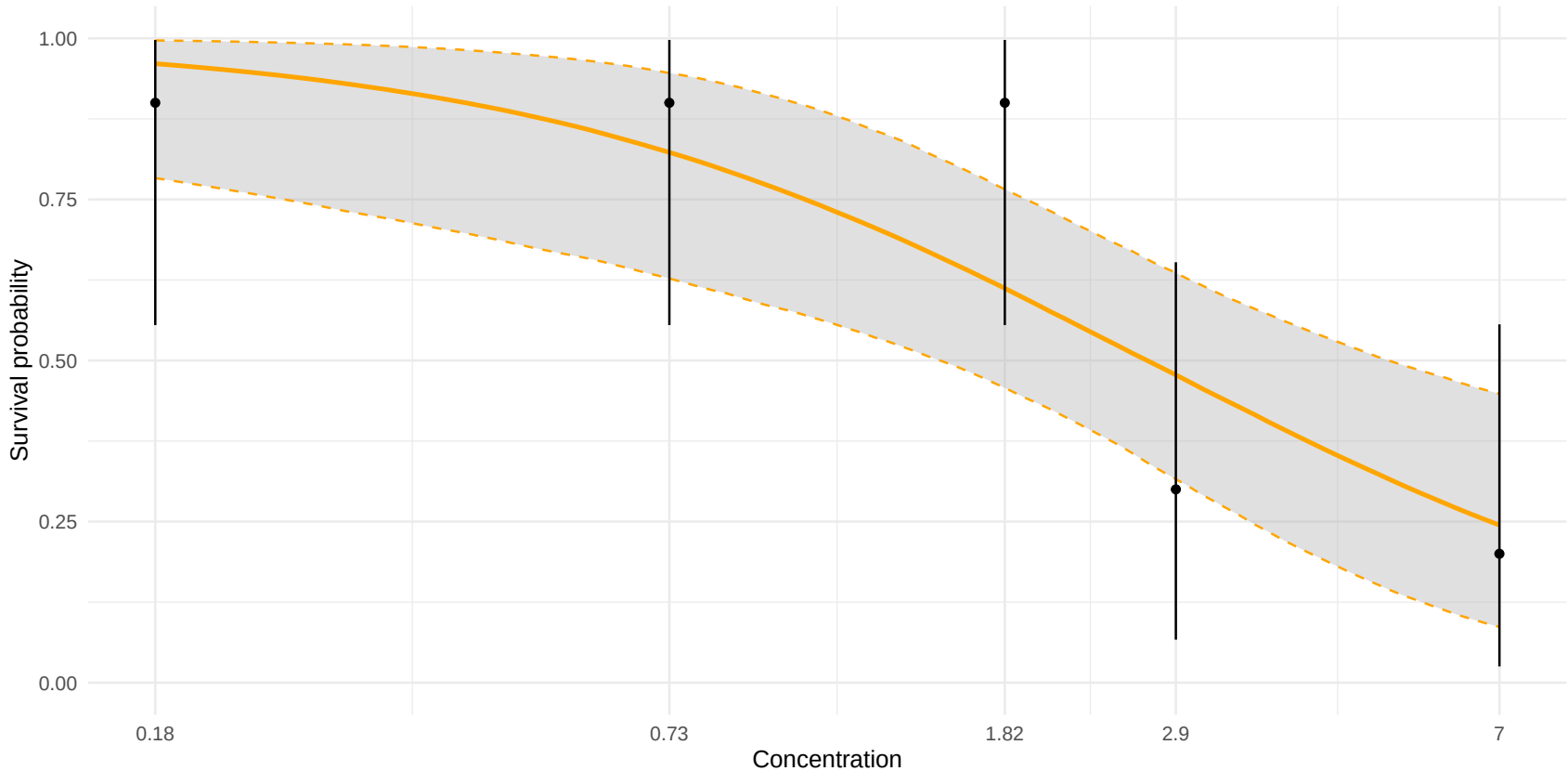
```
             50%       2.5%      97.5%
LC10 4.191e-01 2.786e-02 1.031e+00
LC20 8.295e-01 1.467e-01 1.606e+00
LC50 2.682e+00 1.528e+00 5.175e+00
```

As already stated above, $e = LC_{50}$.

DR model for binary data: fitting plot

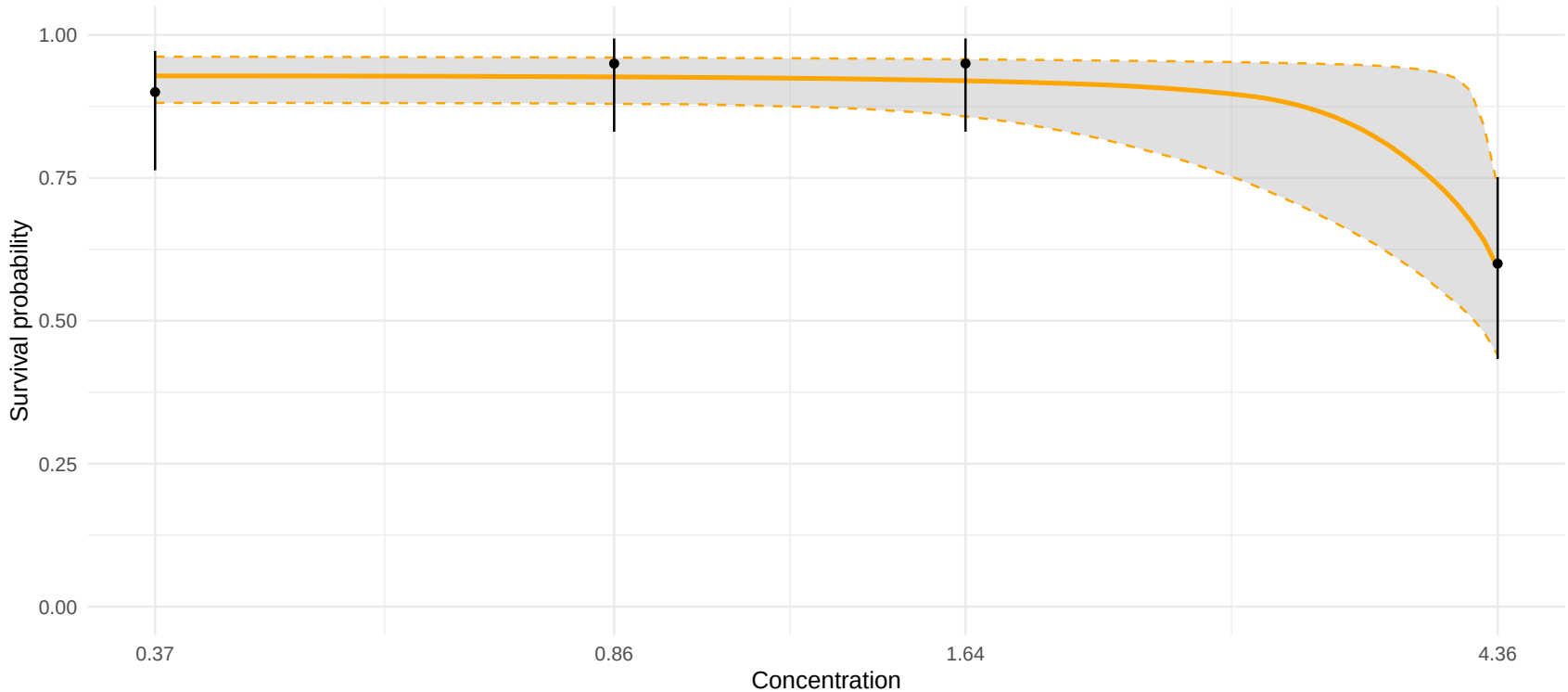
The orange line corresponds to the median predicted DR relationship, while in gray is displayed the 95% credible band. Bot the median curve and its uncertainty band are obtained by propagating uncertainties on parameters into the prediction.

```
plot(fit, adddata = TRUE, log.scale=TRUE)
```



DR model for binary data: possible warnings

```
Warning: The LC50 estimation (model parameter e) lies outsi
tested concentrations and may be unreliable as the prior di
parameter is defined from this range !
```



DR model for binary data: posterior predictive check (definition)

- black dots are observations;
- segments are predictions of the expected numbers of success at each concentration.

We expect to see black dots at points of coordinates $y = x$ (**represented as steps when replicates are shifted on the x-axis**).

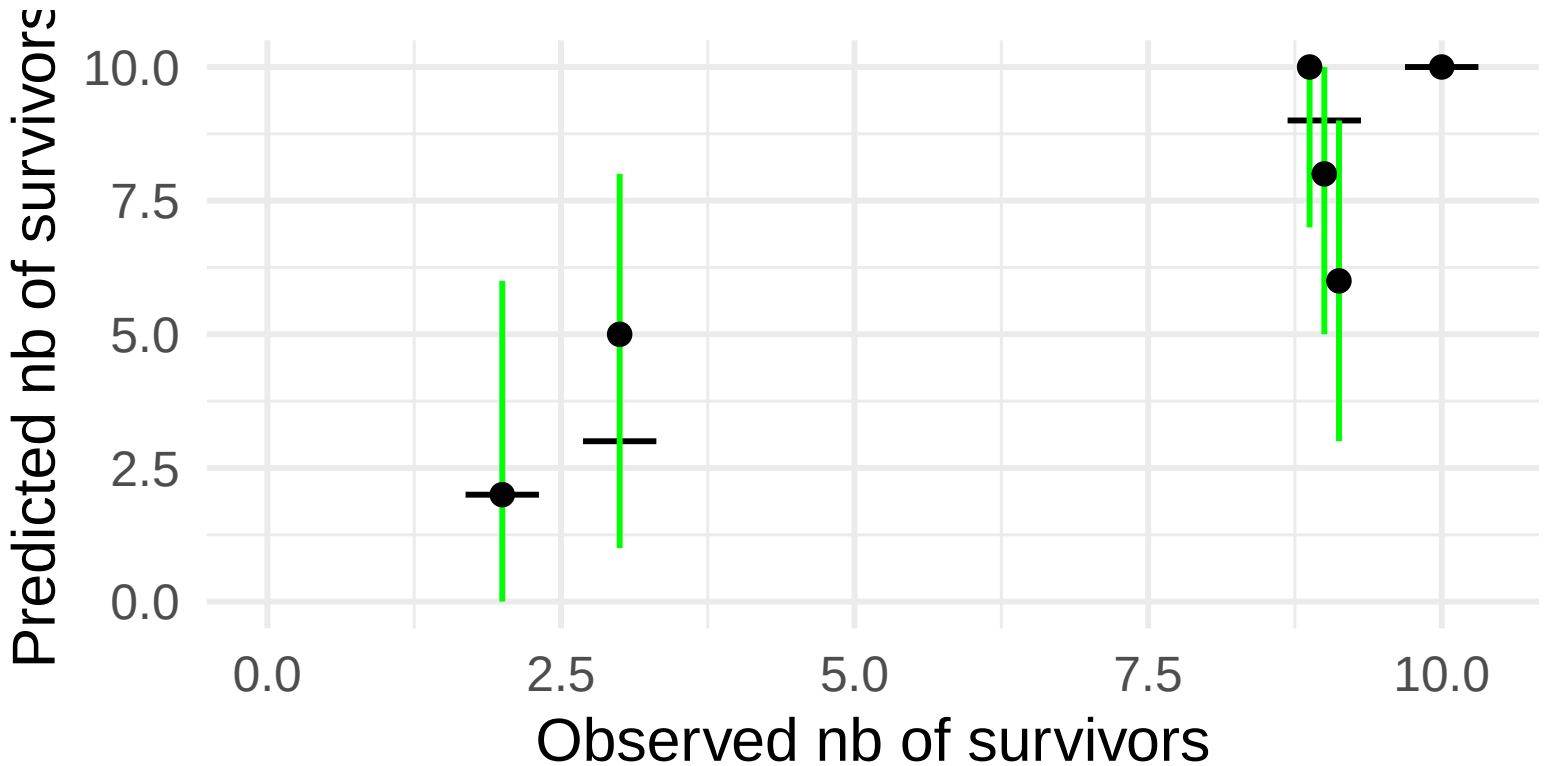
The model provides a reasonable variation around the predicted mean value as an interval where we expect to have 95% of the dots in average.

Intervals are coloured in green if they overlap the line $y = x$, in red otherwise.

DR model for binary data: posterior predictive check (plot)

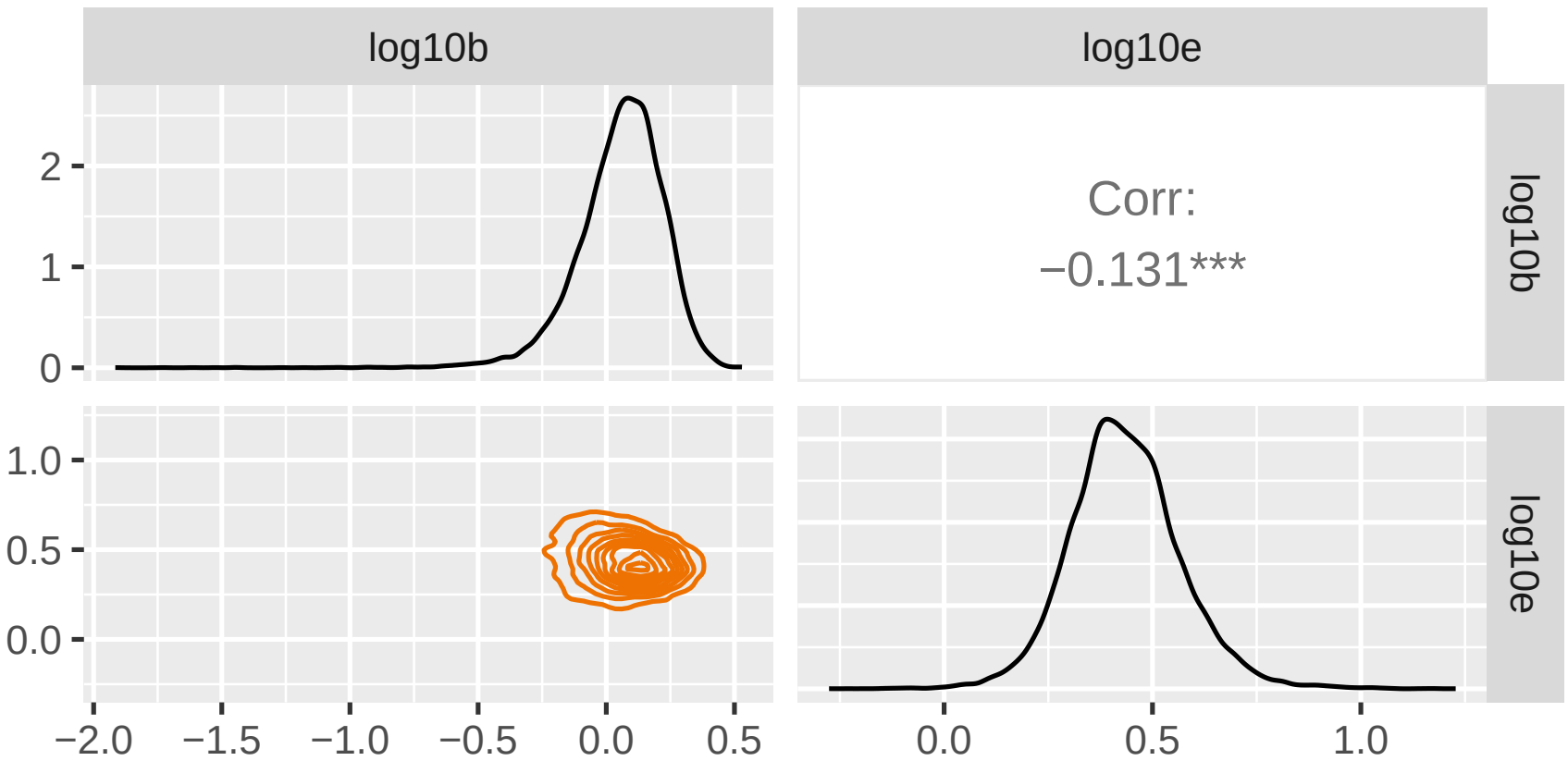
$y = x$ looks like steps because replicates are shifted on the x-axis.

```
ppc(fit)
```



DR model for binary data: parameter correlations

```
ggsmcmc <- ggs(fit$mcmc)
ggs_pairs(ggsmcmc,
          lower = list(continuous = wrap("density",
                                         color = "#ee7202")))
```



Look at morse count data

```
dataset <- reproData(chlordan)
# Create en R object to be used with 'morse'
# --> Add columns 'Ninit', 'Nindtime' and 'Nreprocumul'
dataset[50:56,6:8]
```

```
# A tibble: 7 x 3
  Ninit Nindtime Nreprocumul
  <int>    <dbl>       <int>
1     1        5           0
2     1        6           0
3     1        7           8
4     1        8           8
5     1        9          23
6     1       10          23
7     1       11          23
```

Look at raw count data

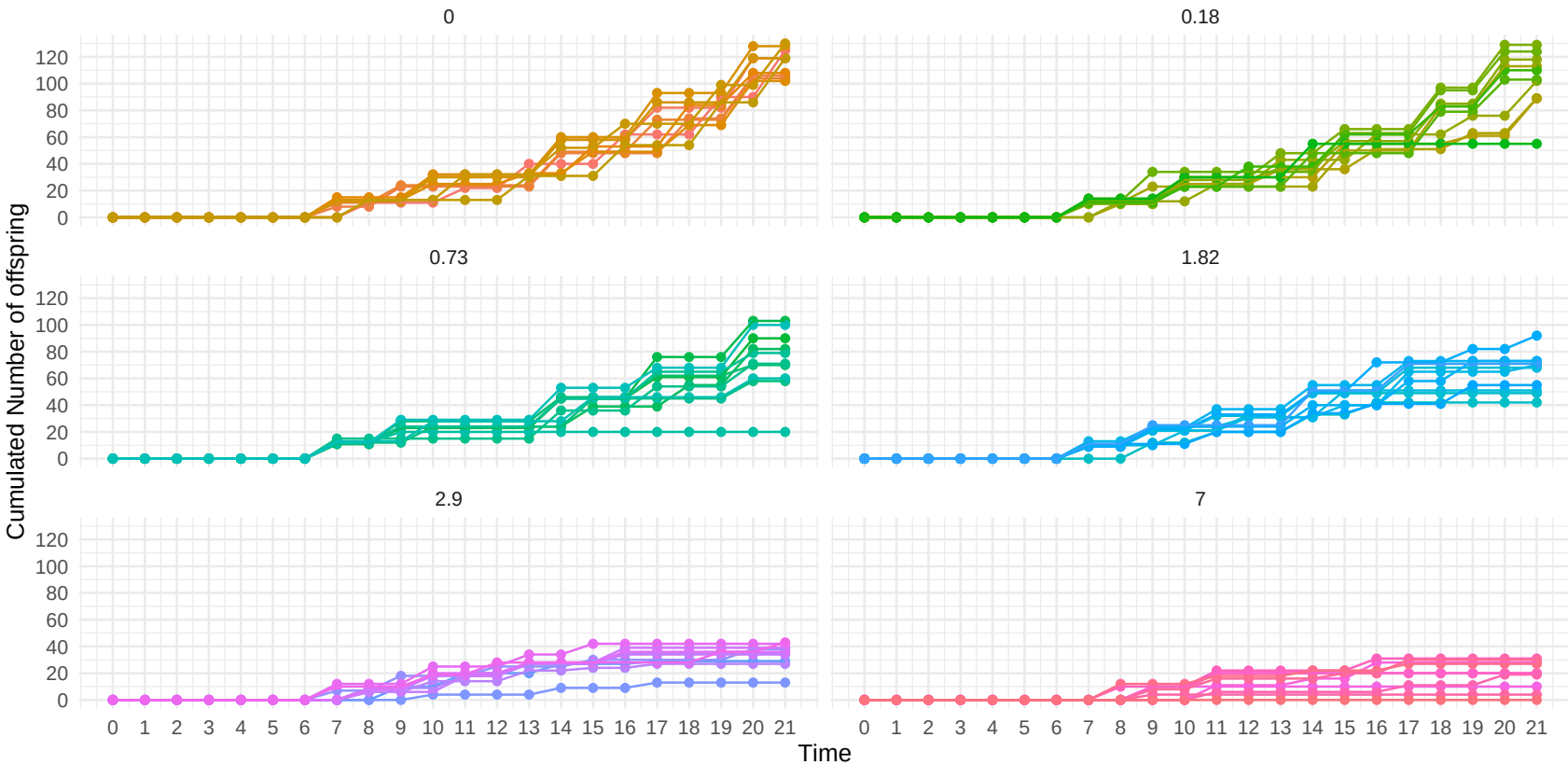
```
# Number of offspring (sum of replicates)
# per time and concentration
summary(dataset, quiet=TRUE)$NboffTimeConc
```

```
     0 1 2 3 4 5 6   7  8   9  10 11 12 13  14  15 16  17
0    0 0 0 0 0 0 0  84 37  28  86 30  0 44 152  37 63 109
0.18 0 0 0 0 0 0 0  97 22  34 105 13 15 56  50 143 34   0 1
0.73 0 0 0 0 0 0 0 121  0  90  27  0  0  0 128  54  0 116
1.82 0 0 0 0 0 0 0  92  0 101  11 53 29  0 137  29 38 100
2.9  0 0 0 0 0 0 0  29 37  21  56 31 16 53  12  15 34   9
7    0 0 0 0 0 0 0   0 22  40   0 65  0  0   9   4 21  20
```

Plot raw count data

Cumulated number of something (e.g., offspring) versus time at each concentration and each replicate.

```
plot(dataset)
```

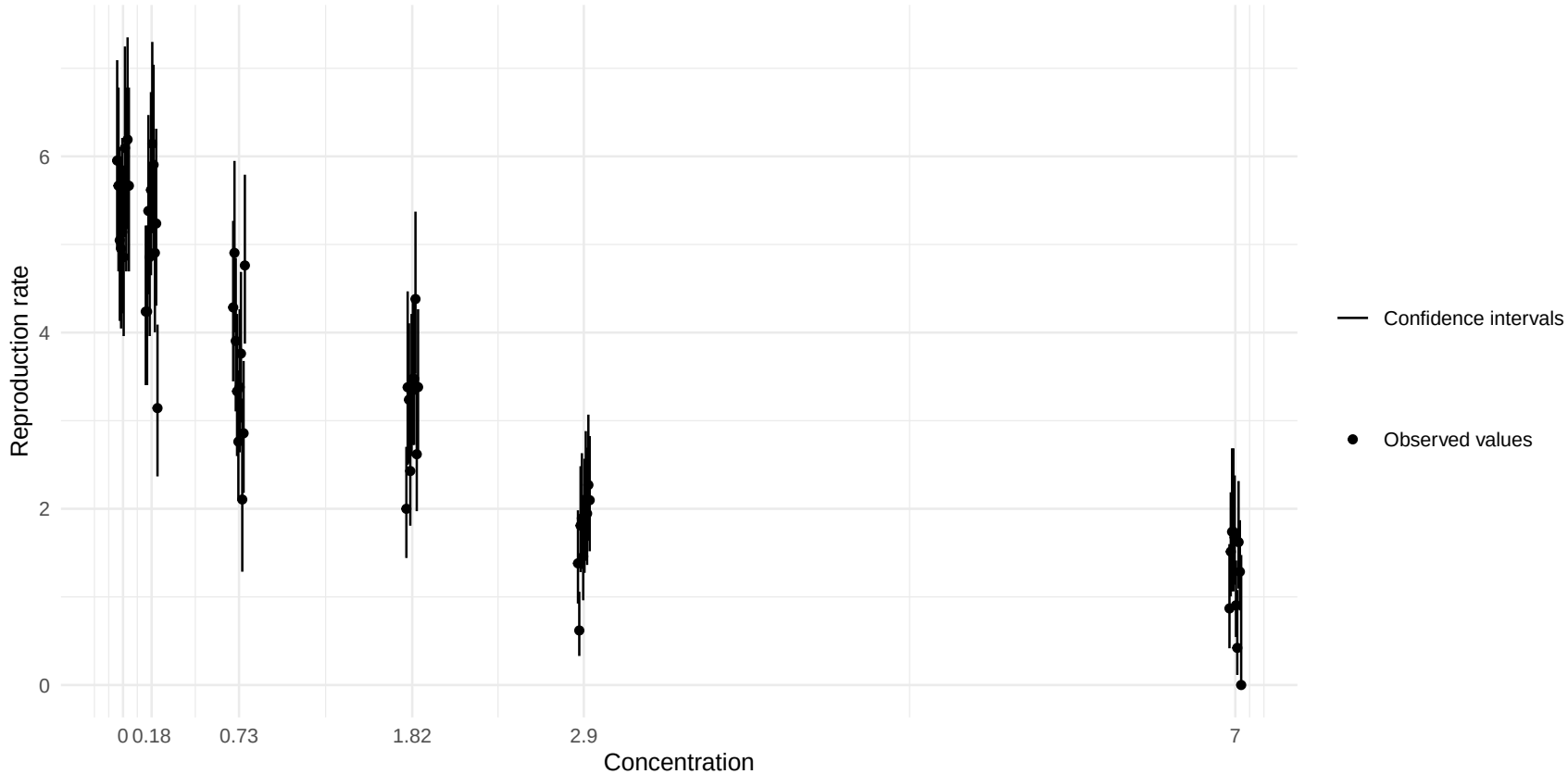


Plot the concentration-effect curve at any target time

For example, reproduction rate versus concentration at day 21 (target time).

Poisson confidence intervals on data (95%)

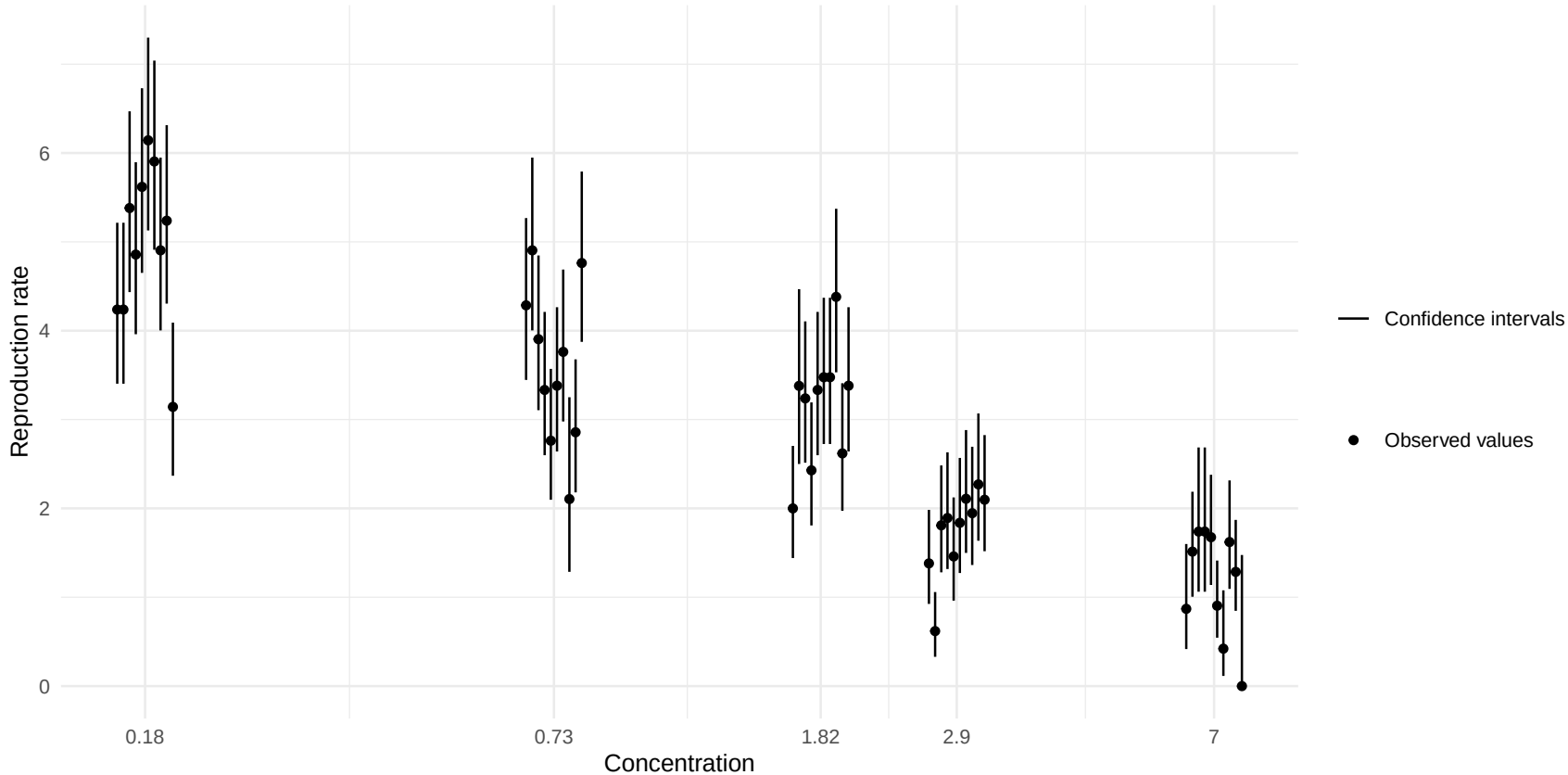
```
plotDoseResponse(dataset, target.time = 21)
```



Plot the concentration-effect curve at any target time

Change the x-scale.

```
plotDoseResponse(dataset, target.time = 21,
                 addlegend = TRUE, log.scale=TRUE)
```



→ Mortality in some replicates & inter-replicate variability.

A co-variable: the number of organism-days (*NID*)

For each organism, we estimate the period it has stayed alive (which we assume coincides with the period it may reproduce).

As commonly done in epidemiology for incidence rate calculations, we then calculate, for one replicate, the total sum of the periods of observation of each organism before its death.

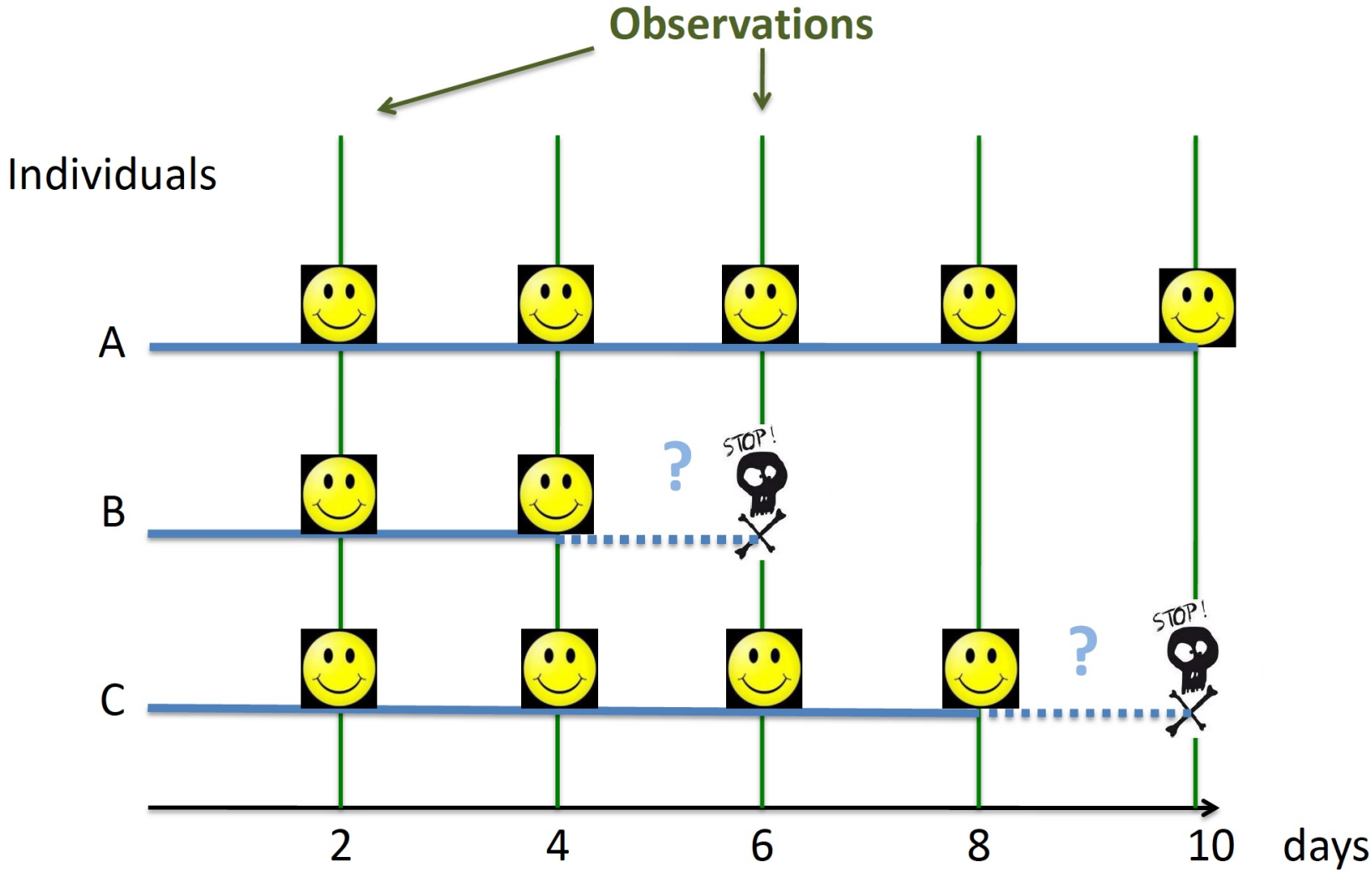
This sum is expressed as the **number of organism-days**.

Hence, reproduction is evaluated through the number of offspring per organism-day.

[Delignette-Muller et al., 2014]

A co-variable: the number of organism-days (*NID*)

Below, after 10 days of exposition,
$NID = 10 + \left(\frac{6+4}{2}\right) + \left(\frac{10+8}{2}\right) = 24$ organism-days.



DR model for count data

Within `morse`, the number of counts N_{ij} at c_i in replicate j is modelled using a Poisson distribution:

$$N_{ij} \sim Poisson(f(c_i;\theta) \times NID_{ij})$$

with $f(c_i;\theta) = \frac{d}{1+(\frac{c_i}{e})^b}$. Here d corresponds to what is counted under control conditions and e to the c_i leading a count equal to $\frac{d}{2}$. Parameter b reflects the effect intensity of the contaminant. Is the variability between replicates cannot be neglected, then:

$$N_{ij} \sim Poisson(f_{ij} \times NID_{ij})$$

where f_{ij} at c_i in replicate j is become a random variable following a Gamma distribution with an **over-dispersion** parameter ω:

$$f_{ij} \sim gamma(\frac{f(c_i;\theta)}{\omega}, \frac{1}{\omega})$$

$\rightarrow$ the greater ω, the greater the inter-replicate variability.

DR model for count data: priors

Posterior distributions for parameters b, d, e and ω are estimated using JAGS with the following priors:

$$\log_{10} e \sim \mathcal{N}(\mu_e, \sigma_e)$$

with $\mu_e = \frac{\log_{10}(\min c_i)+\log_{10}(\max c_i)}{2}$

and $\sigma_e = \frac{\log_{10}(\max c_i)-\log_{10}(\min c_i)}{4}$

$$d \sim \mathcal{N}(\mu_d, \sigma_d)$$

with $\mu_d = \frac{1}{r_0}\sqrt{\sum_j \frac{N_{0j}}{NID_{0j}}}$

and $\sigma_d = \sqrt{\frac{\sum_j \frac{N_{0j}}{NID_{0j}} - \mu_d}{r_0(r_0-1)}}$

$$\log_{10} b \sim \mathcal{U}(-2, 2) \text{ et } \log_{10} \omega \sim \mathcal{U}(-4, 4)$$

Poisson or gamma-Poisson?

For a given data set, the procedure implemented in `morse` fit both models and the Deviance Information Criterion (DIC) is used to choose the most appropriate one.

In situations where over-dispersion is negligible, using the Poisson model will provide more reliable estimates. Hence, a Poisson model is preferred unless the Gamma-Poisson model has a sufficiently lower DIC (in practice we require a difference of 10).

DR analysis for count data: fitting at day 21

```
fit <- reproFitTT(dataset, quiet=TRUE,
                  target.time = 21,
                  ecx = c(10, 20, 50))
```

Here the target time is fixed to 21 days (by default, then the end of the experiment).

We can ask for the estimation of any EC_x values:

here $x = 10, 20, 50\%$.

DR analysis for count data: parameter estimates

Get parameter estimates as medians and 95% credible intervals:

```
summary(fit, quiet=TRUE)$Qpost
```

```
              50%       2.5%      97.5%
b      9.253e-01 7.197e-01 1.176e+00
d      5.520e+00 5.233e+00 5.815e+00
e      1.772e+00 1.376e+00 2.250e+00
omega 1.526e-01 8.287e-02 2.739e-01
```

What do you notice?

DR analysis for count data: priors-posteriors

```
summary(fit, quiet=TRUE)$Qpriors[,2:3]
```

```
             2.5%     97.5%
b      1.259e-02 7.943e+01
d      5.220e+00 5.827e+00
e      7.467e-01 6.843e+00
omega 1.585e-04 6.310e+03
```

```
summary(fit, quiet=TRUE)$Qpost[,2:3]
```

```
             2.5%     97.5%
b      7.197e-01 1.176e+00
d      5.233e+00 5.815e+00
e      1.376e+00 2.250e+00
omega 8.287e-02 2.739e-01
```

DR analysis for count data: EC_x estimates

Get EC_x estimates as medians and 95% credible intervals

```
summary(fit, quiet=TRUE)$QECx
```

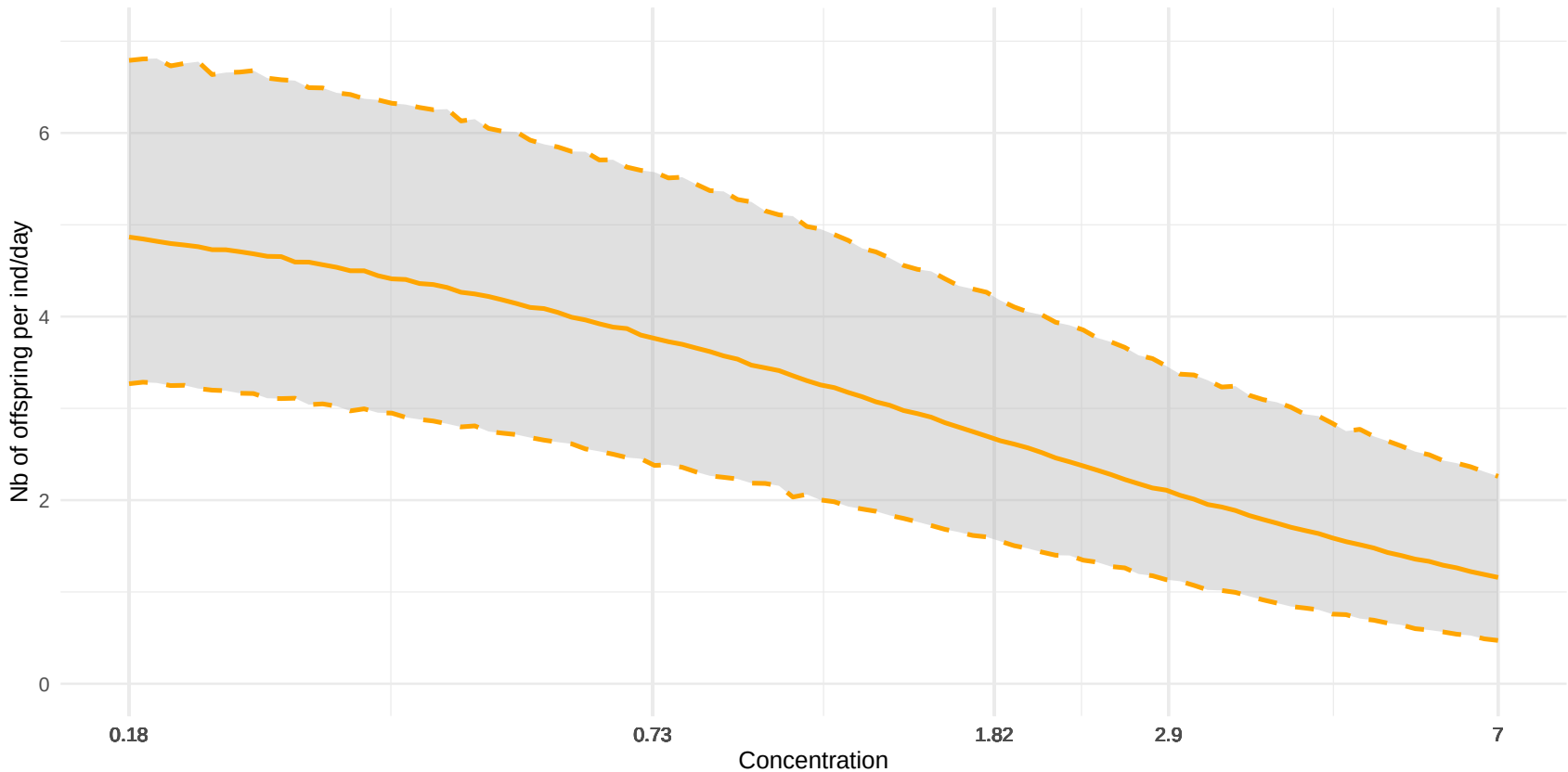
```
             50%       2.5%      97.5%
EC10 1.651e-01 7.345e-02 3.132e-01
EC20 3.960e-01 2.227e-01 6.299e-01
EC50 1.772e+00 1.376e+00 2.250e+00
```

As already defined, $e = EC_{50}$.

DR analysis for count data: fitting plot

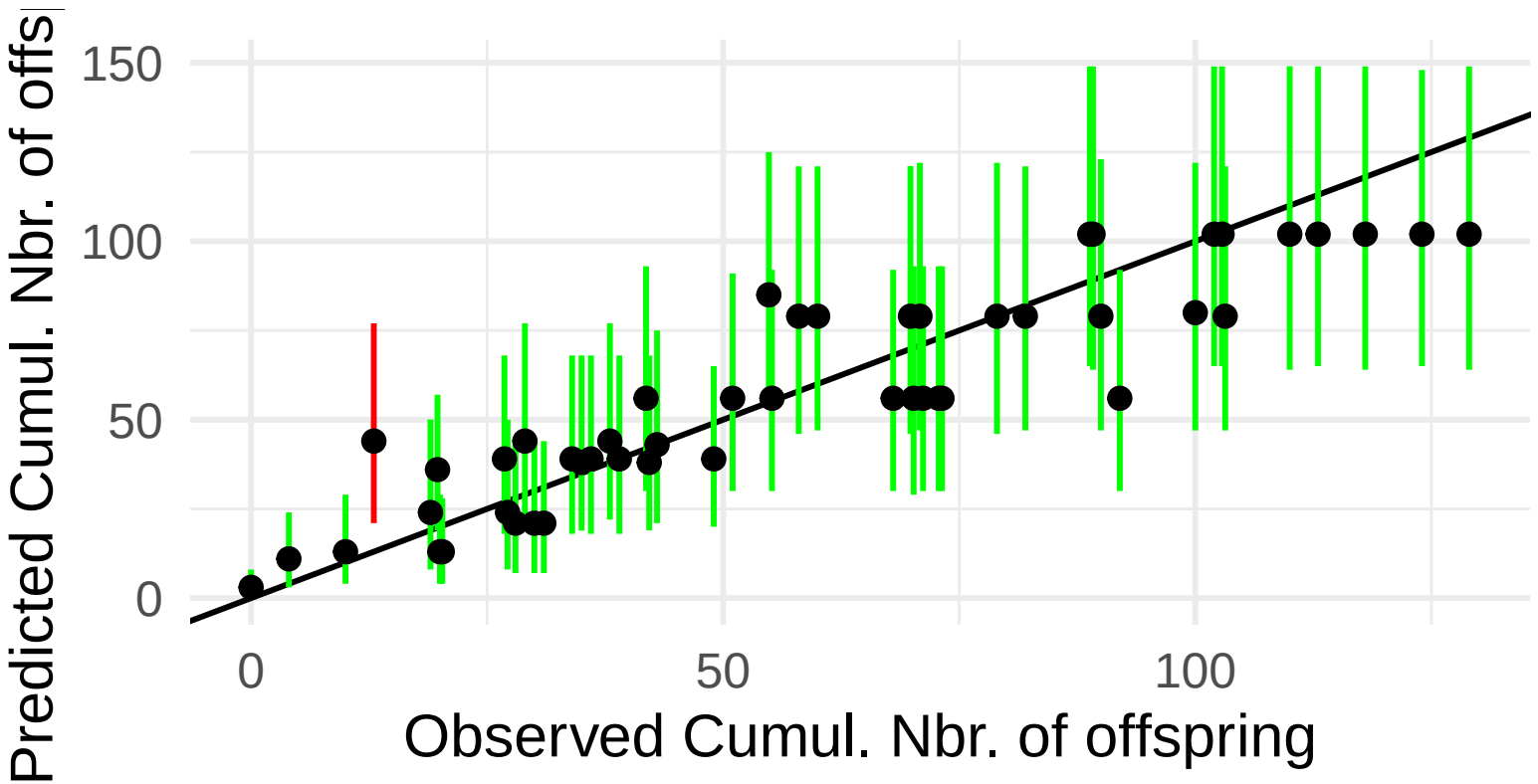
The orange line corresponds to the median predicted DR relationship, while in gray is displayed the 95% credible band. Bot the median curve and its uncertainty band are obtained by propagating uncertainties on parameters into the prediction.

```
plot(fit, log.scale=TRUE)
```



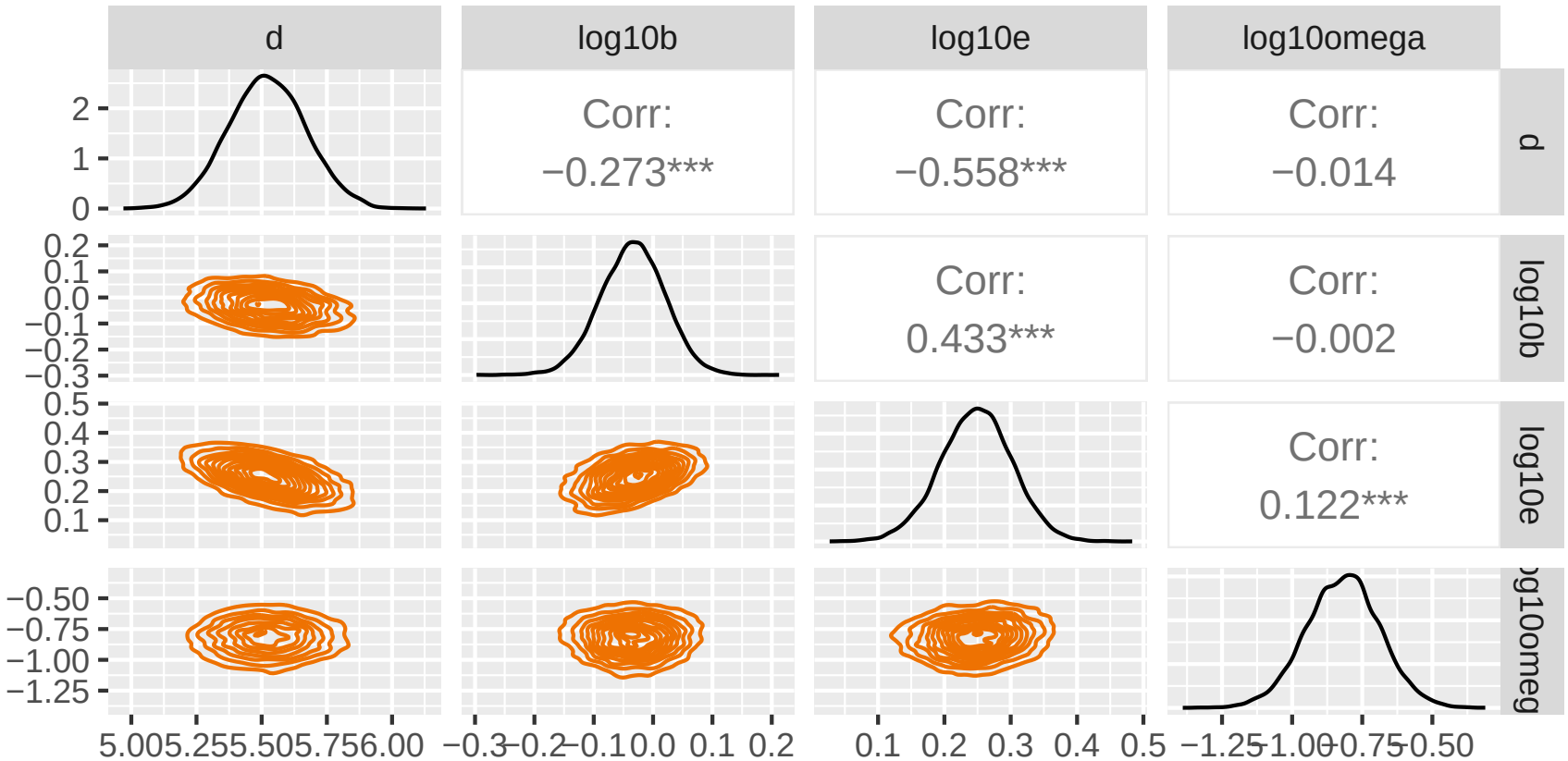
DR analysis for count data: posterior predictive check (plot)

```
ppc(fit)
```



DR analysis for count data: parameter correlations

```
ggsmcmc <- ggs(fit$mcmc)
ggs_pairs(ggsmcmc,
          lower = list(continuous = wrap("density",
                                         color = "#ee7202")))
```



Practice

Go to the practical guide

and

Do it yourself!