Dose-response modelling - Bayesian Inference

The R-package *morse* for survival and reproduction data

Professor Sandrine CHARLES

2022-09-27

Contact: sandrine.charles@univ-lyon1.fr

About 'morse'

Current version: 3.3.1

Baudrot, V., Charles, S., 2021. morse: an R-package in support of Environmental Risk Assessment. *Journal of Open Source Software* 6, 3200. https://doi.org/10.21105/joss.03200

https://cran.r-project.org/src/contrib/Archive/morse/morse_3.3.1.tar.gz

[On-line; accessed 04-February-2021]

Get an example data set

Reproduction and survival 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. Reproduction and survival 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 survival 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 survival 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 survival data (3)

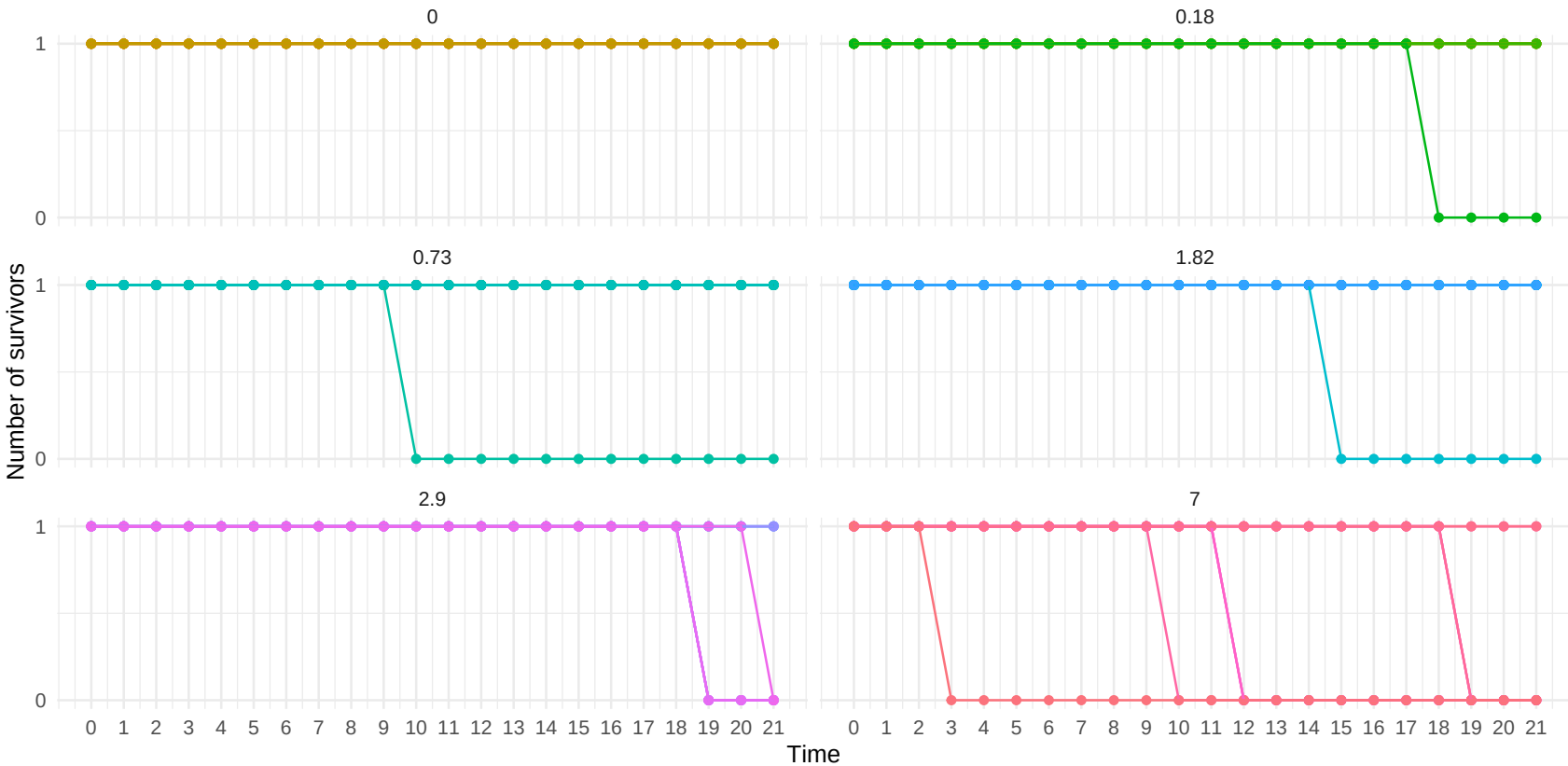
```
# Number of survivors (sum of replicates)
# per time and concentration
summary(dataset, quiet=TRUE)$NbsurvTimeConc
```

```
      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 survival data

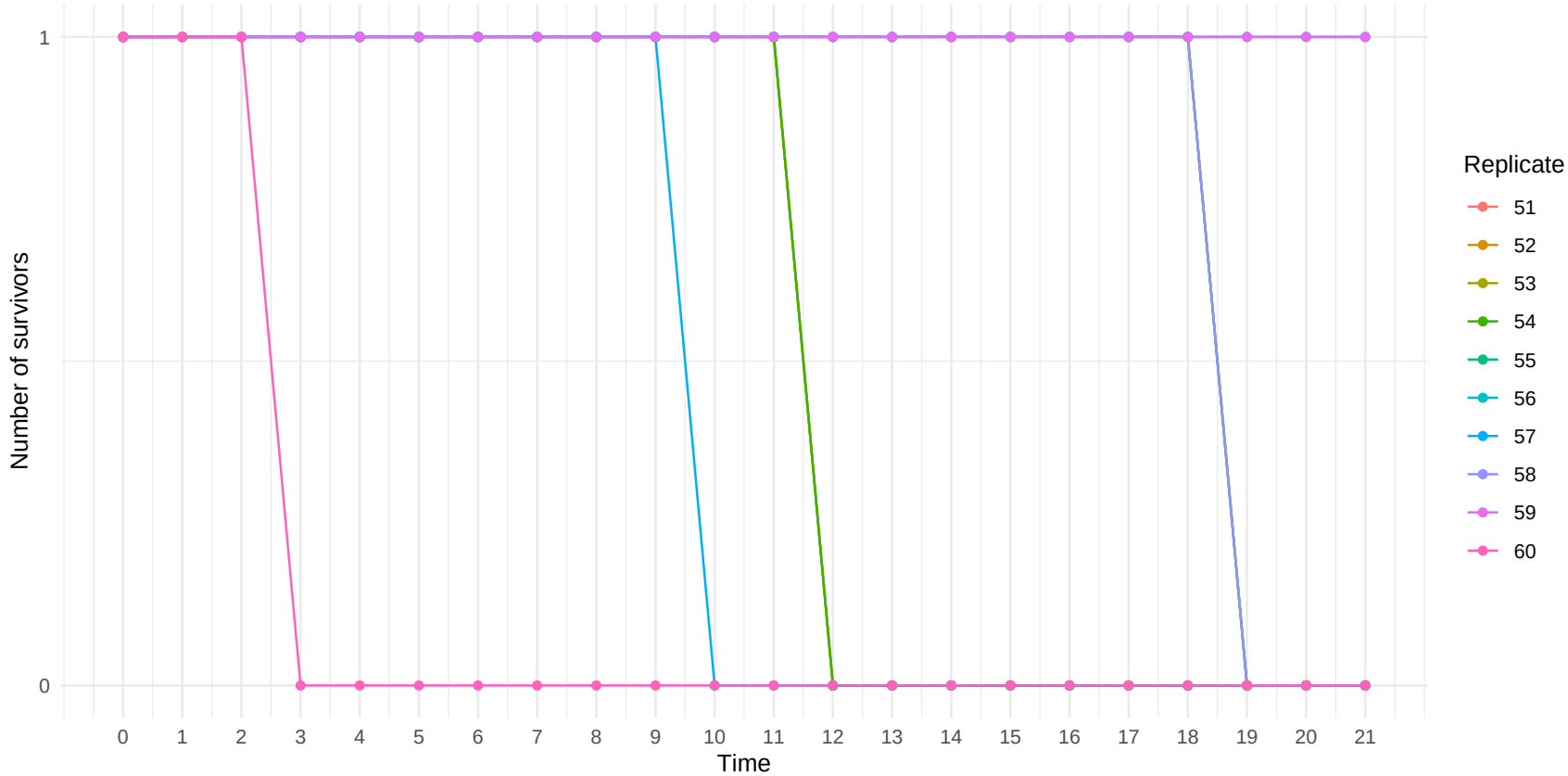
Number of surviving individuals versus time at each concentration and each replicate.

```
plot(dataset)
```



Focus on one concentration in particular

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

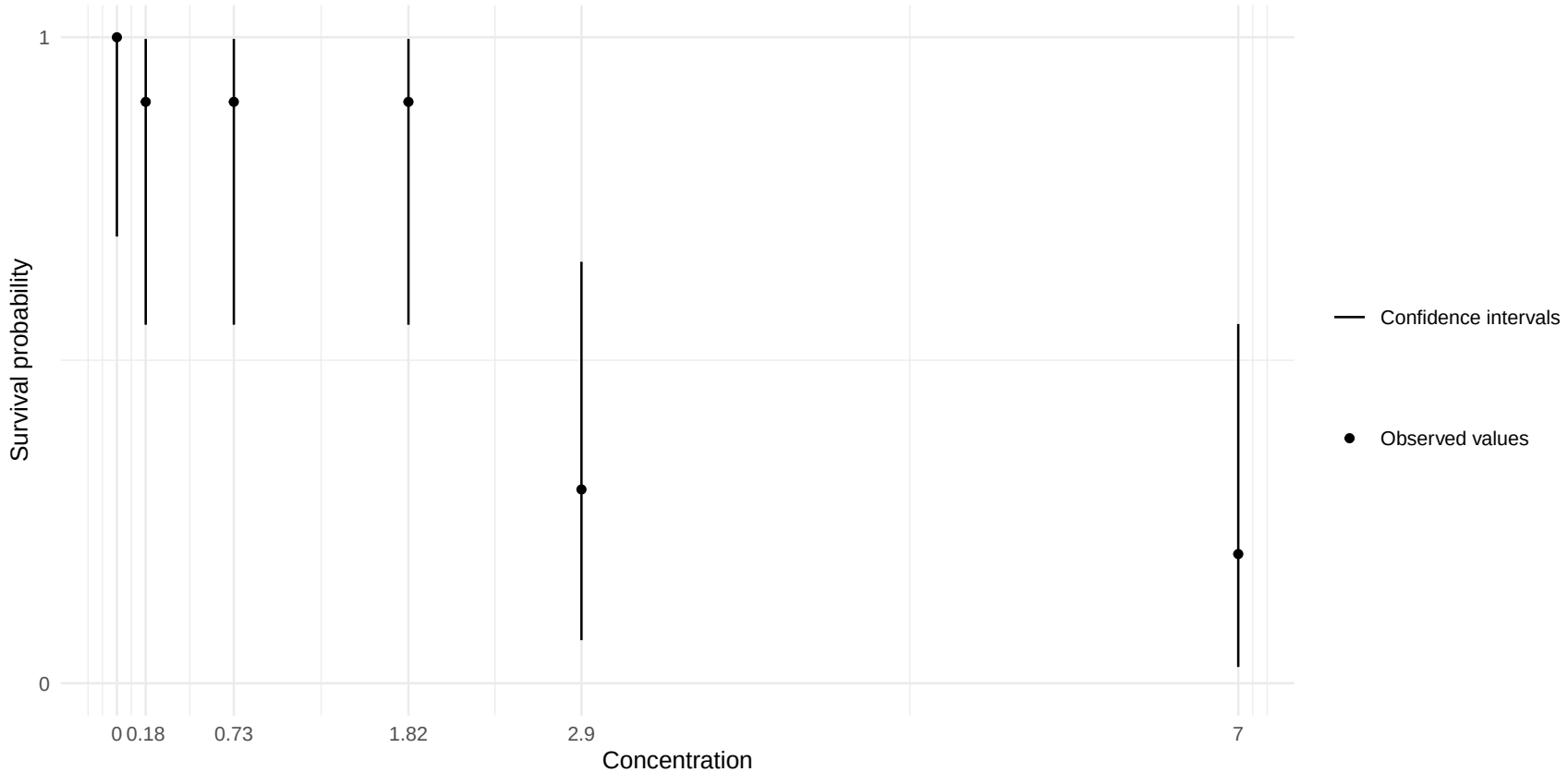


Plot the concentration-response curve at any target time

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

Binomial confidence intervals are added to the observed values.

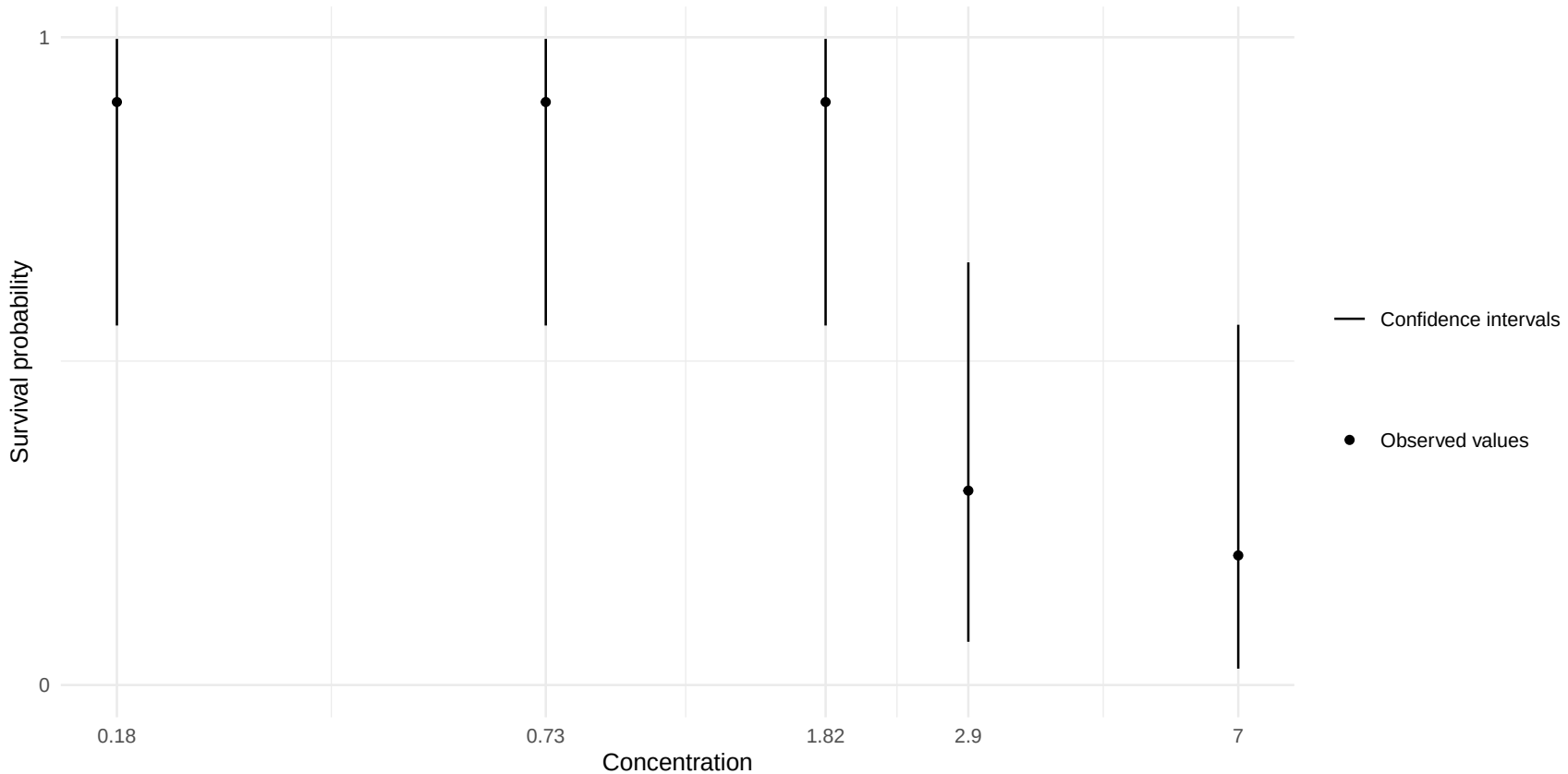
```
plotDoseResponse(dataset, target.time = 21,
                 addlegend = TRUE)
```



Plot the concentration-response curve at any target time

Change the x-scale

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



Survival model: a reminder

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

Survival data are quantal data, so that the number N_i of surviving individuals 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 individuals 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 while b, e and d are (positive) model parameters: d corresponds to the survival rate in absence of pollutant, e corresponds to the LC_{50} and b (the *slope*) is related to the effect intensity of the contaminant.

Survival model: priors

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

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

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

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

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

If there is no mortality in control experiments, then we fix $d = 1$, otherwise we assume $d \sim \mathcal{U}(0, 1)$.

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

Survival analysis: 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 $x = 10, 20, 50\%$.

Survival analysis: parameter estimates

Get parameter estimates as medians and 95% credible intervals:

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

```
        50%      2.5%     97.5%
b 1.175e+00 4.84e-01 2.147e+00
e 2.693e+00 1.52e+00 5.323e+00
```

What do you notice?

Survival analysis: 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 4.84e-01 2.147e+00
e 1.52e+00 5.323e+00
```

Survival analysis: LC_x estimates

Get LC_x estimates as medians and 95% credible intervals

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

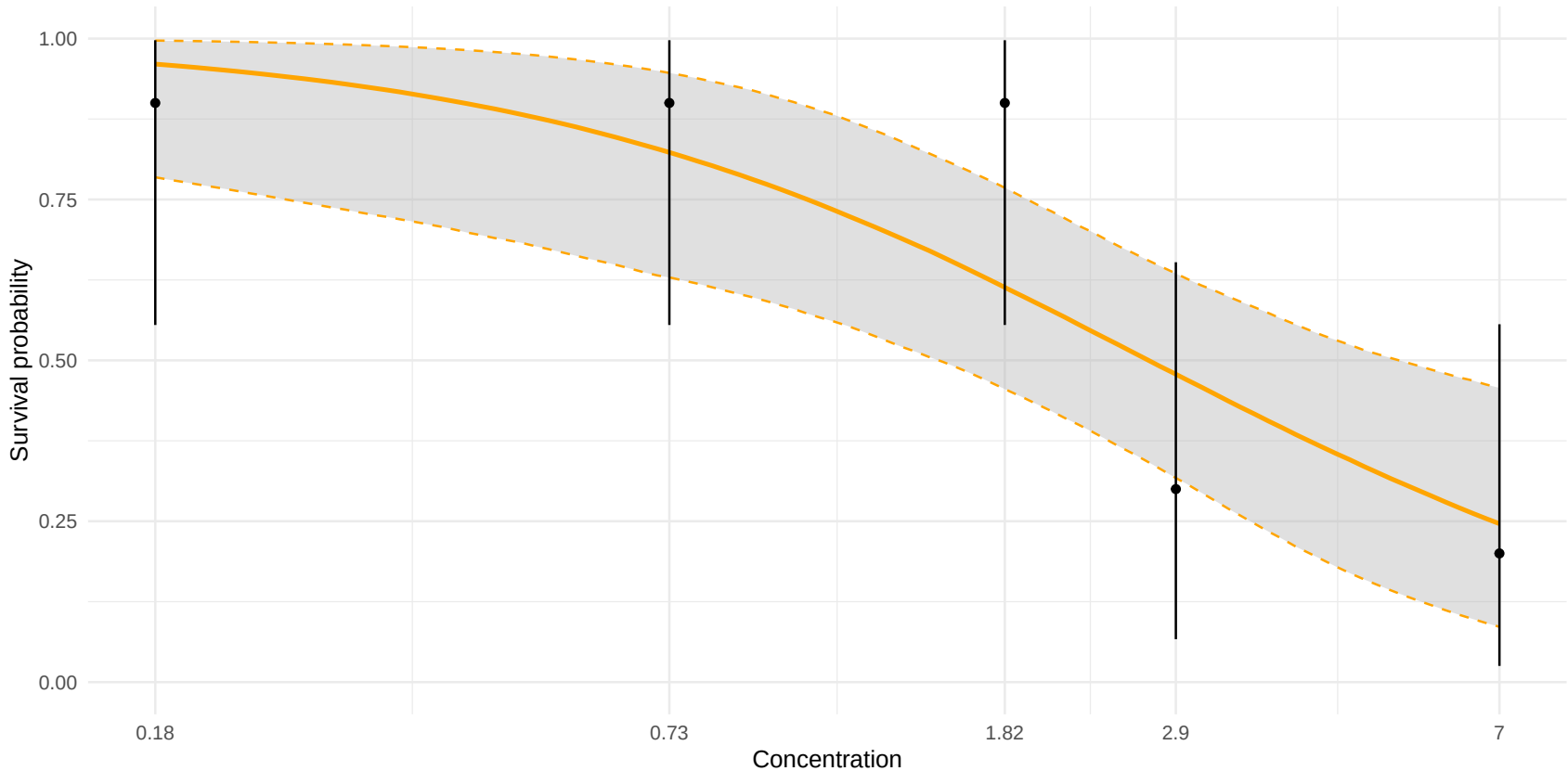
```
             50%       2.5%      97.5%
LC10 4.167e-01 3.068e-02 1.032e+00
LC20 8.305e-01 1.486e-01 1.615e+00
LC50 2.693e+00 1.520e+00 5.323e+00
```

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

Survival analysis: fitting plot

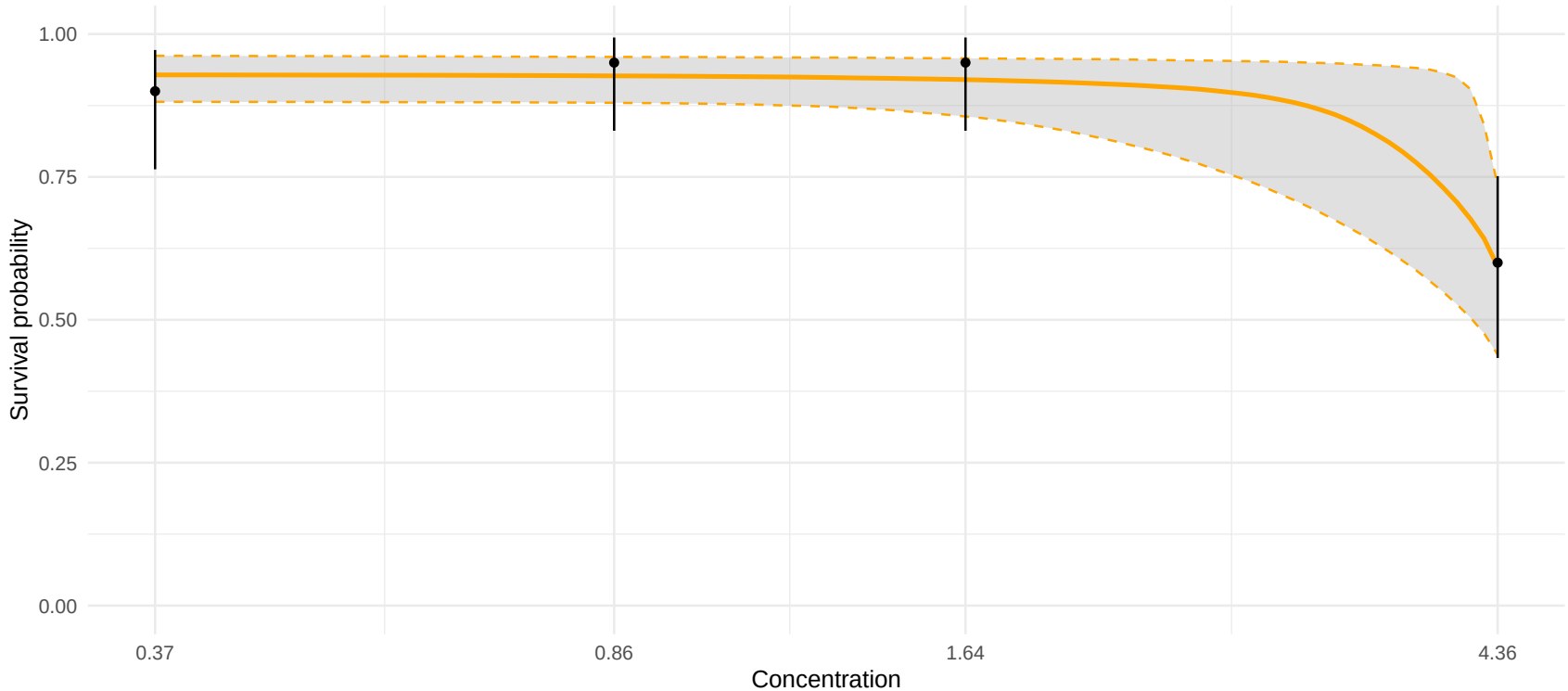
Each parameter is equal to the median value of its posterior (orange curve). In gray, the 95% credible band (propagated uncertainties).

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



Survival analysis: 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 !
```



Survival analysis: posterior predictive check (definition)

- black dots are observations;
- segments are predictions of the expected numbers of survivors 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**).

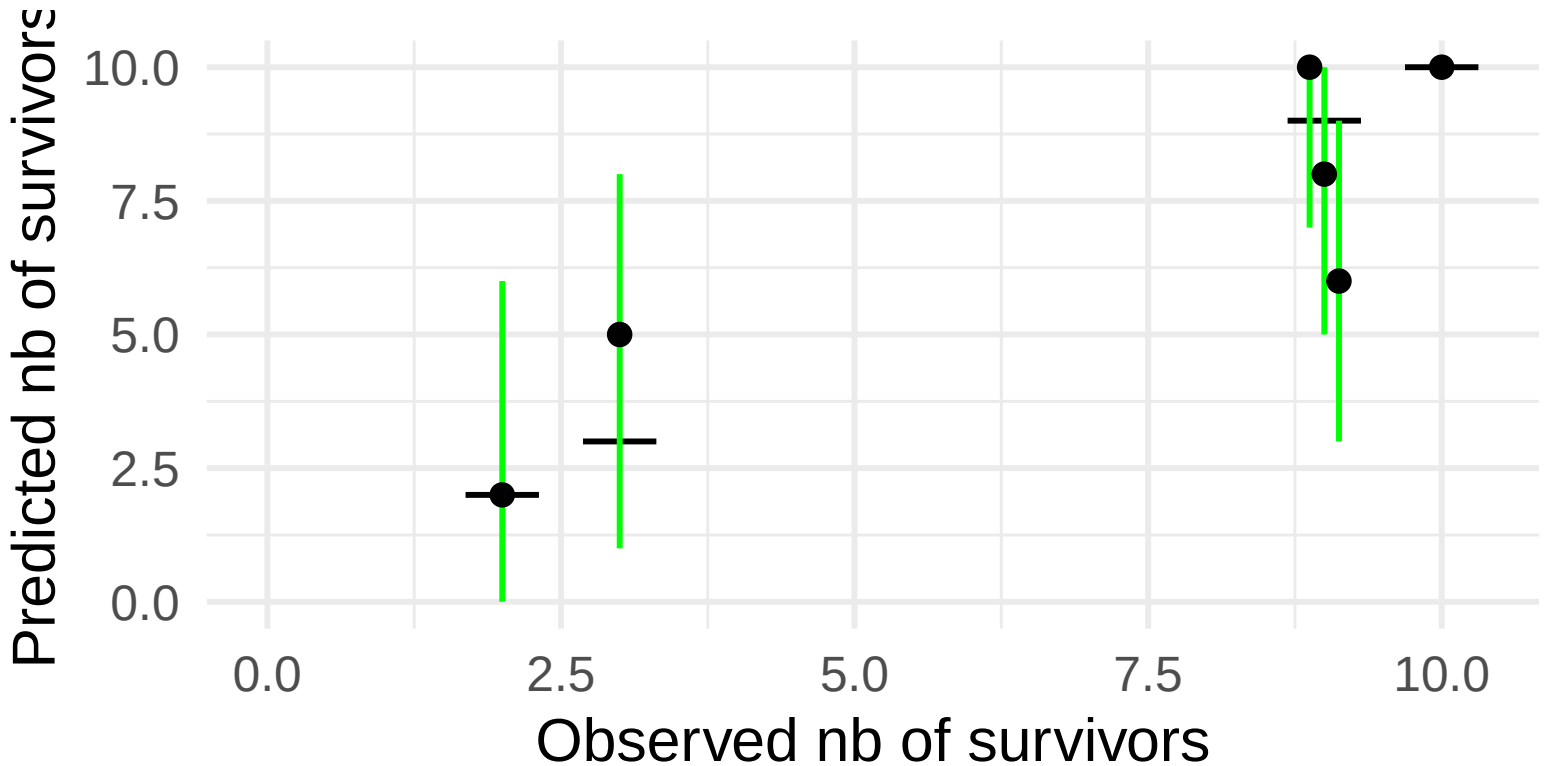
The model provides a tolerable variation around the predited mean value as an interval where we expect 95% of the dots to be in average.

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

Survival analysis: posterior predictive check (plot)

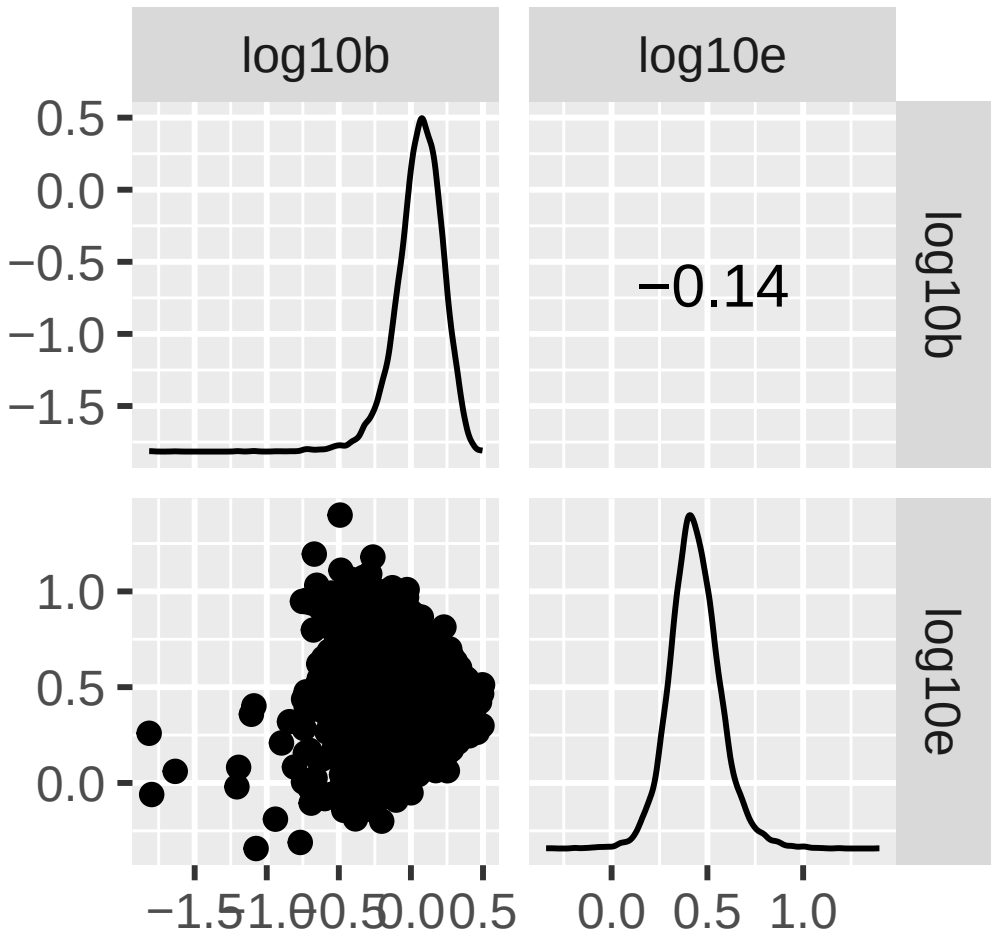
$y = x$ looks like steps because replicates are shifted on the x-axis.

```
ppc(fit)
```



Survival analysis: parameter correlations

```
fit.mcmctot <- as.data.frame(as.matrix(fit$mcmc))
library(GGally)
ggscatmat(fit.mcmctot)
```



Look at 'morse' reproduction 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 the raw reproduction data

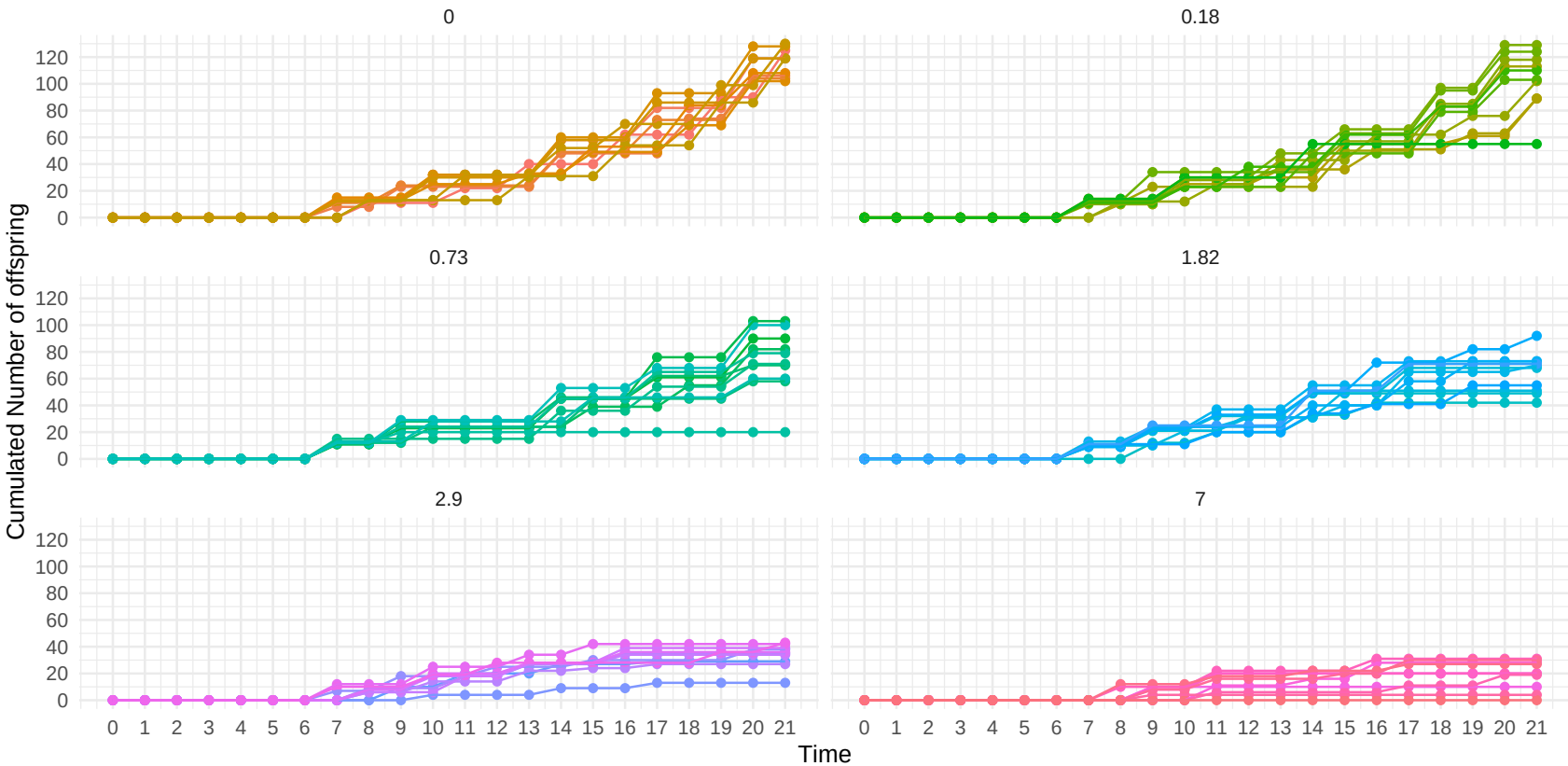
```
# Number of offsprings (sum of replicates)
# per time and concentration
summary(dataset, quiet=TRUE)$NboffTimeConc
```

```
     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 reproduction data

Cumulated number of offsprings versus time at each concentration and each replicate.

```
plot(dataset)
```

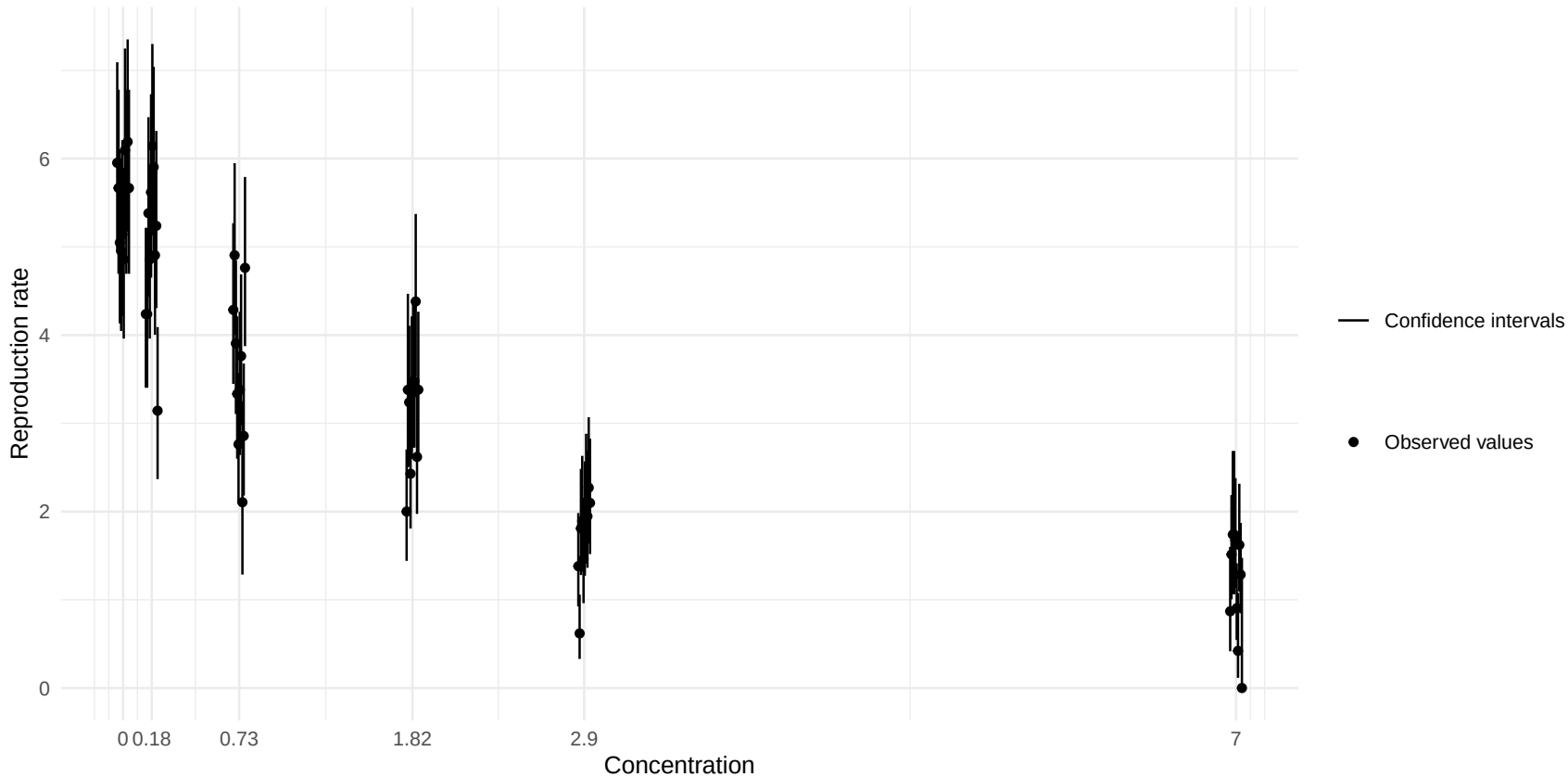


Plot the concentration-effect curve at any target time

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

Poisson confidence intervals on data (95%)

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

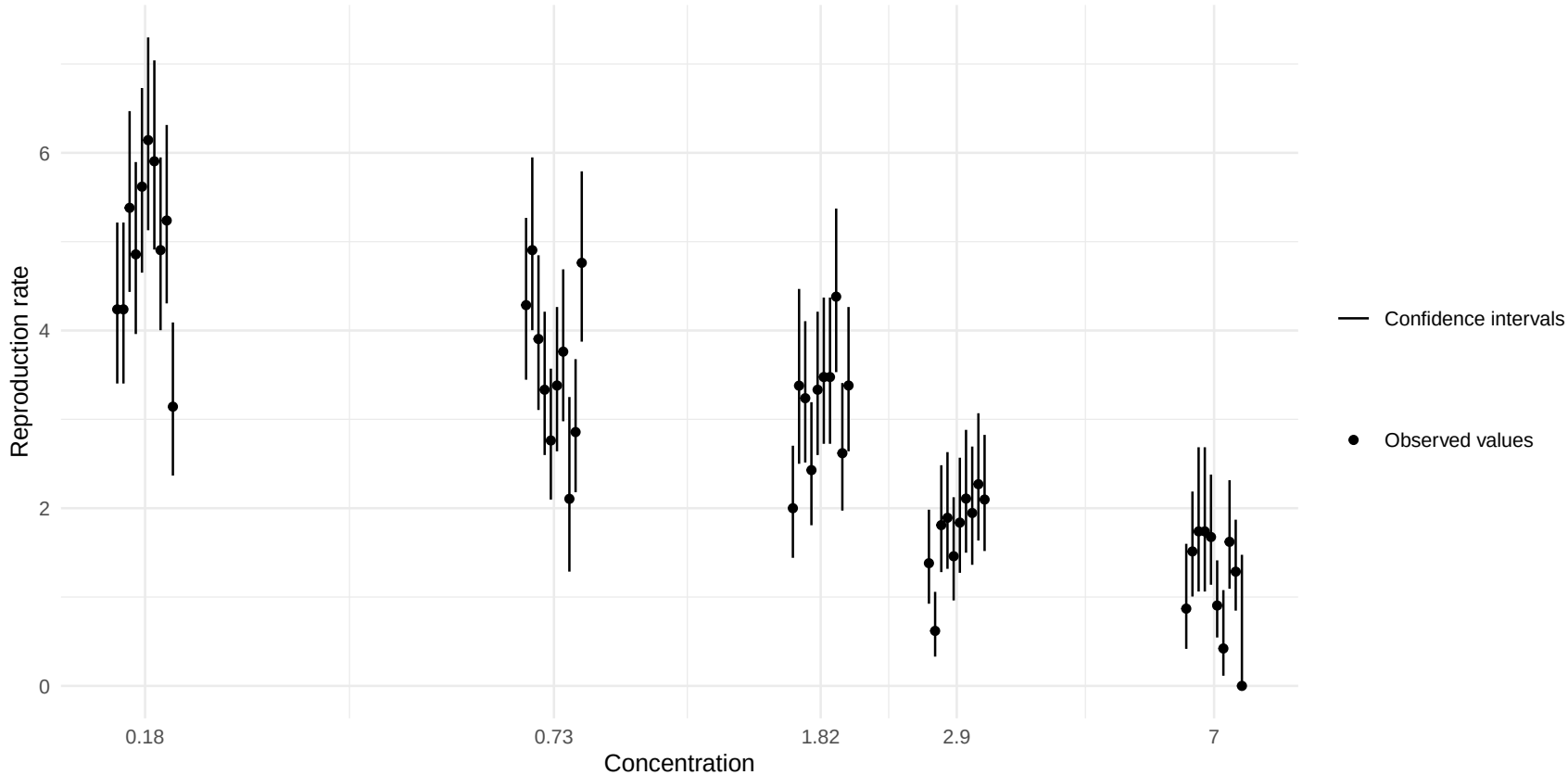


Plot the concentration-effect curve at any target time

Change the x-scale.

$\rightarrow$ Mortality in some replicates, inter-replicate variability.

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



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

We estimate for each individual 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 can then calculate, for one replicate, the total sum of the periods of observation of each individual before its death.

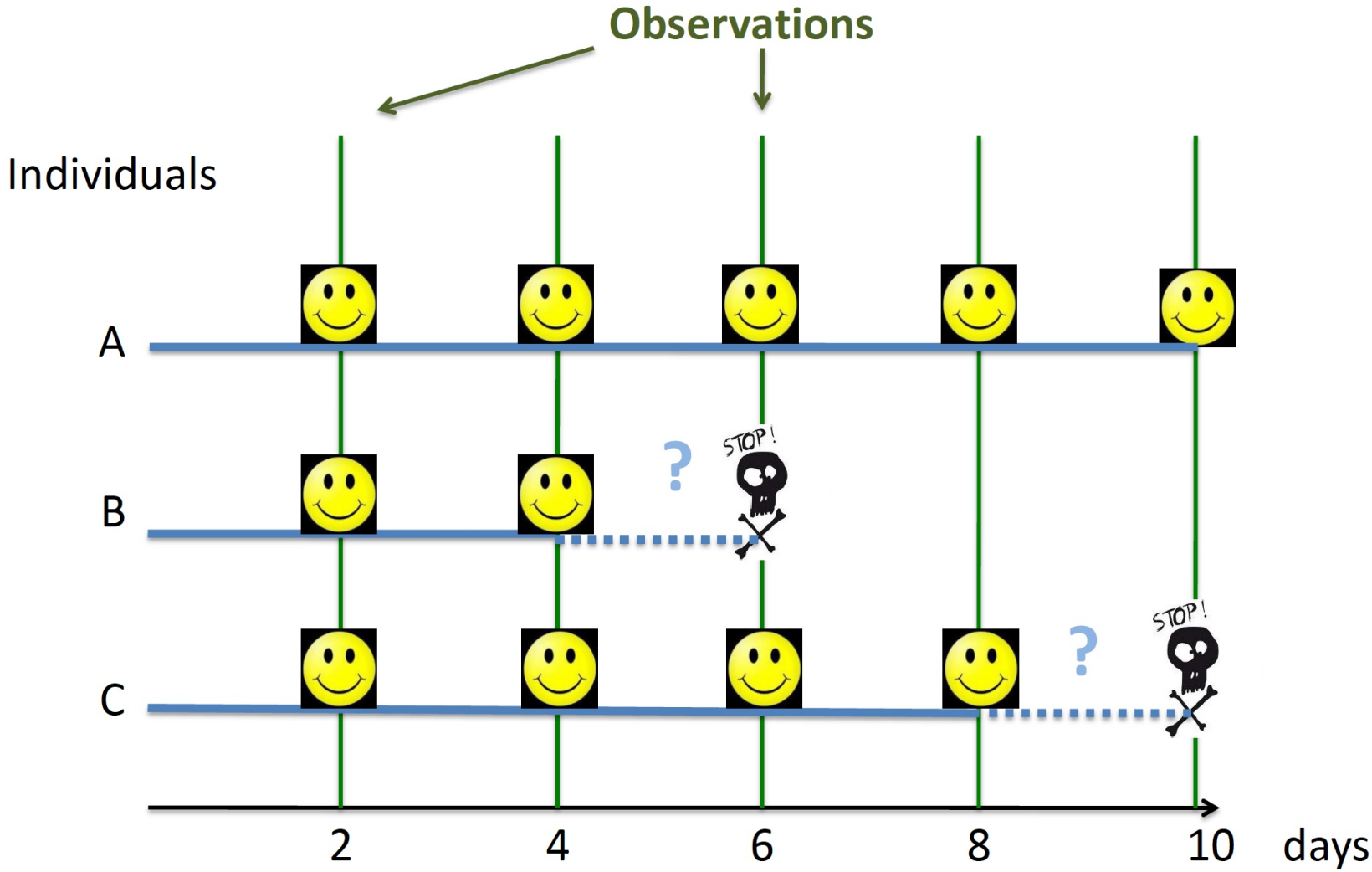
This sum is expressed as the number of individual-days.

Hence, reproduction is evaluated through the number of offsprings per individual-day.

[Delignette-Muller et al., 2014]

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

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



Reproduction model

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

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

avec $f(c_i;\theta) = \frac{d}{1+(\frac{c_i}{e})^b}$. Here d corresponds to the control reproduction rate and e to the value of the EC_{50}. Parameter b reflects the effect intensity of the contaminant.

In case of a non-negligible variability of the reproduction rate between replicates for a fixed concentration, then:

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

where reproduction rate f_{ij} at c_i in replicate j is a random variable following a gamma distribution with a dispersion parameter ω:

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

ω is an overdispersion parameter: the greater its value, the greater the inter-replicate variability.

Reproduction model: priors

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

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

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

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

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

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

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

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

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.

In situations where overdispersion 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).

Reproduction analysis: 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 (default).

We can ask for the estimation of any EC_x values:

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

Reproduction analysis: parameter estimates

Get parameter estimates as medians and 95% credible intervals:

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

```
              50%       2.5%      97.5%
b     9.243e-01 7.148e-01 1.184e+00
d     5.520e+00 5.227e+00 5.810e+00
e     1.770e+00 1.373e+00 2.261e+00
omega 1.521e-01 8.409e-02 2.697e-01
```

What do you notice?

Reproduction analysis: 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     7.148e-01 1.184e+00
d     5.227e+00 5.810e+00
e     1.373e+00 2.261e+00
omega 8.409e-02 2.697e-01
```

Reproduction analysis: EC_x estimates

Get EC_x estimates as medians and 95% credible intervals

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

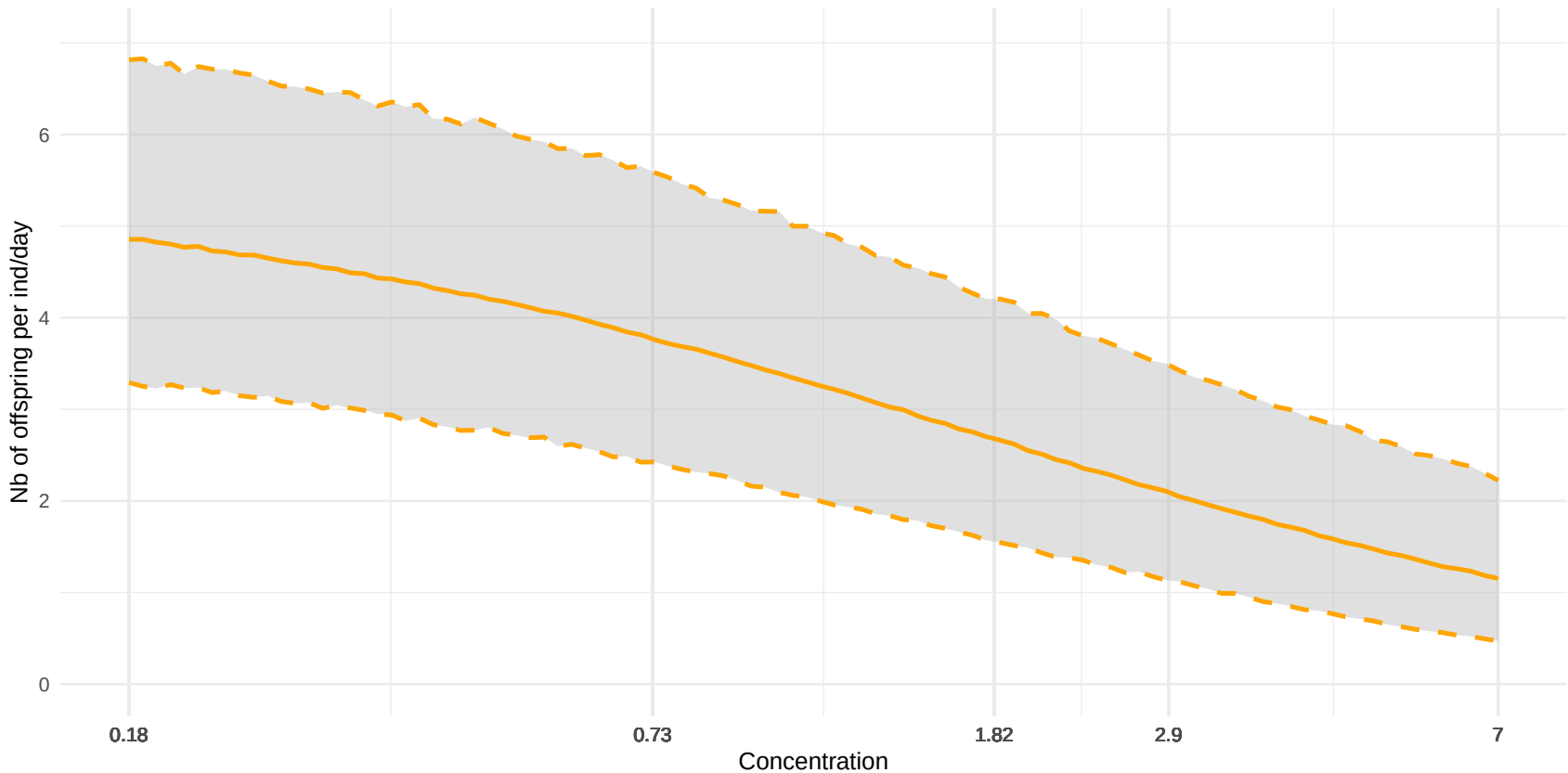
```
             50%       2.5%      97.5%
EC10 1.640e-01 7.253e-02 3.177e-01
EC20 3.938e-01 2.203e-01 6.399e-01
EC50 1.770e+00 1.373e+00 2.261e+00
```

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

Reproduction analysis: plot fitting

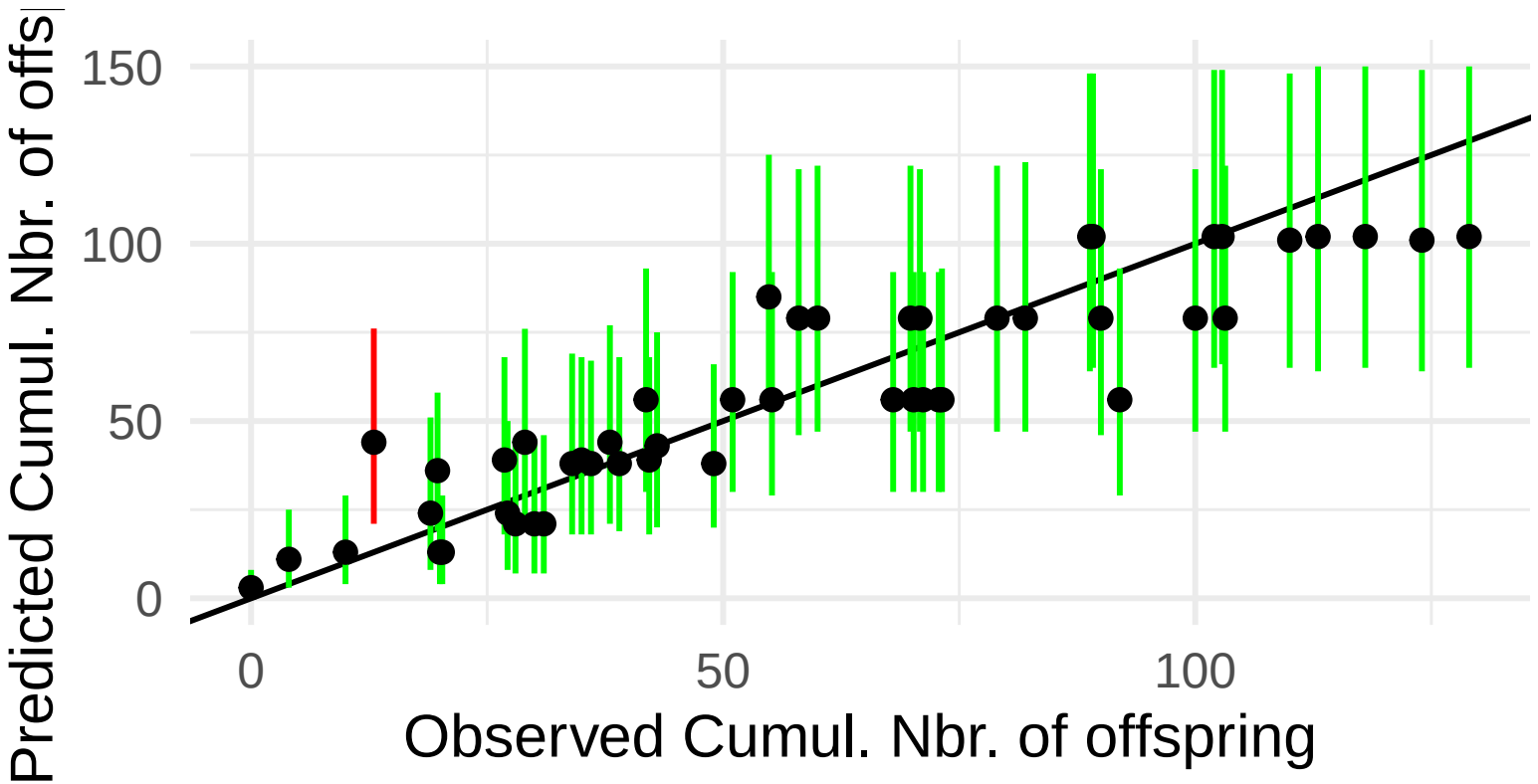
Each parameter is equal to the median value of its posterior (orange curve). In gray, the 95% credible band (propagated uncertainties).

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



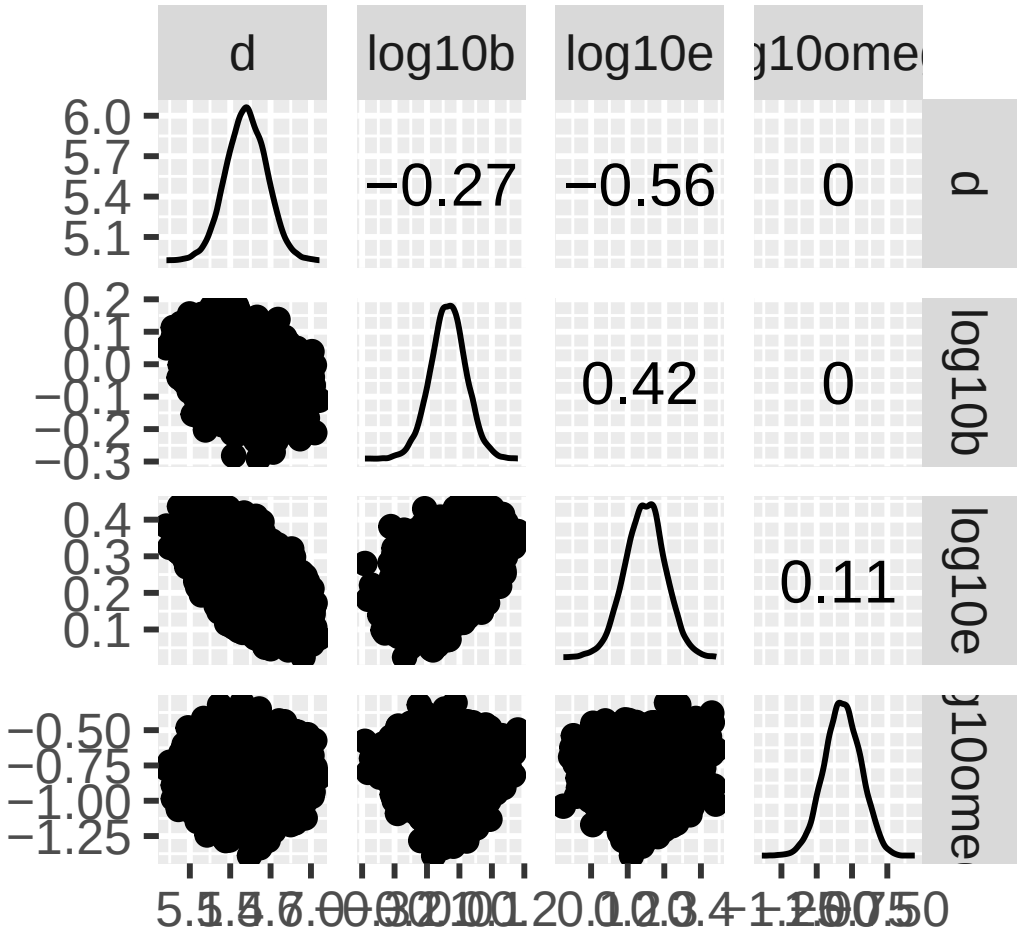
Reproduction analysis: posterior predictive check (plot)

```
ppc(fit)
```



Reproduction analysis: parameter correlations

```
fit.mcmctot <- as.data.frame(as.matrix(fit$mcmc))
ggscatmat(fit.mcmctot)
```



Practice

Go to the practical guide

and

Do it yourself!