

CatReg
Concentration-Time- Response
Modeling of Categorical Data for Risk
Assessment

Jeffrey Gift, Ph.D.
Angela Howard, Ph.D.
National Center for Environmental Assessment
www.epa.gov/ncea

Prepared for

U.S. Environmental Protection Agency
National Center for Environmental Assessment
Research Triangle Park, NC

by



Kenneth G. Brown, Ph.D.*

KBinc@mindspring.com

with

EPA advisors Jeff Gift, Ph.D. and Angela Howard, Ph.D.

*purchase order EPO7C000072

CatReg was created and developed for U.S. EPA by

Douglas G. Simpson, Ph.D, University of Illinois

Minge Xie, Ph.D., Rutgers University

Raymond J. Carroll, Ph.D., Texas A&M University

Henrik Schmiediche, Ph.D., Texas A&M University

Limin Fu, University of Illinois

Additions/modifications:

Kenneth G. Brown, Ph.D., ChaeHyung Ahn, Ph.D.,

Gyoungjin Ahn

Where to obtain CatReg

- Currently, send e-mail to Dr. Jeff Gift at gift.jeff@epa.gov
- In future, it will be available from the EPA NCEA website at www.epa.gov/ncea under “Tools and Databases”

Risk assessment

- **Components** (generally accepted)
 - hazard identification
 - dose-response assessment
 - exposure assessment
 - risk characterization
- **Not included**
 - policy or regulations
 - risk management
- **Class objective** – show how CatReg can be used to extend the dose-response component to dose-time-response of one or more studies, and obtain PODs for one or more exposure durations

Outline

- Introduction/Overview
- Basic features of CatReg
- Using CatReg

Response data

Severity response (ordered categorical): Two or more naturally ordered severity levels, e.g., none, mild, moderate, severe. Case of exactly two levels is usually called **dichotomous response (quantal, binary)** e.g., nonadverse, adverse

Continuous response: Effects measured on a continuum, e.g., weight, enzyme levels

This presentation: CatReg was developed for severity response data, but can be used with continuous data

Dose-response assessment

Objective: Determine a reference dose or concentration (RfD or RfC)

General procedure:

- Determine the critical effect
- Establish a dose or concentration to use as a point of departure (POD)
- Determine RfC from POD/UF (UF is uncertainty factor)

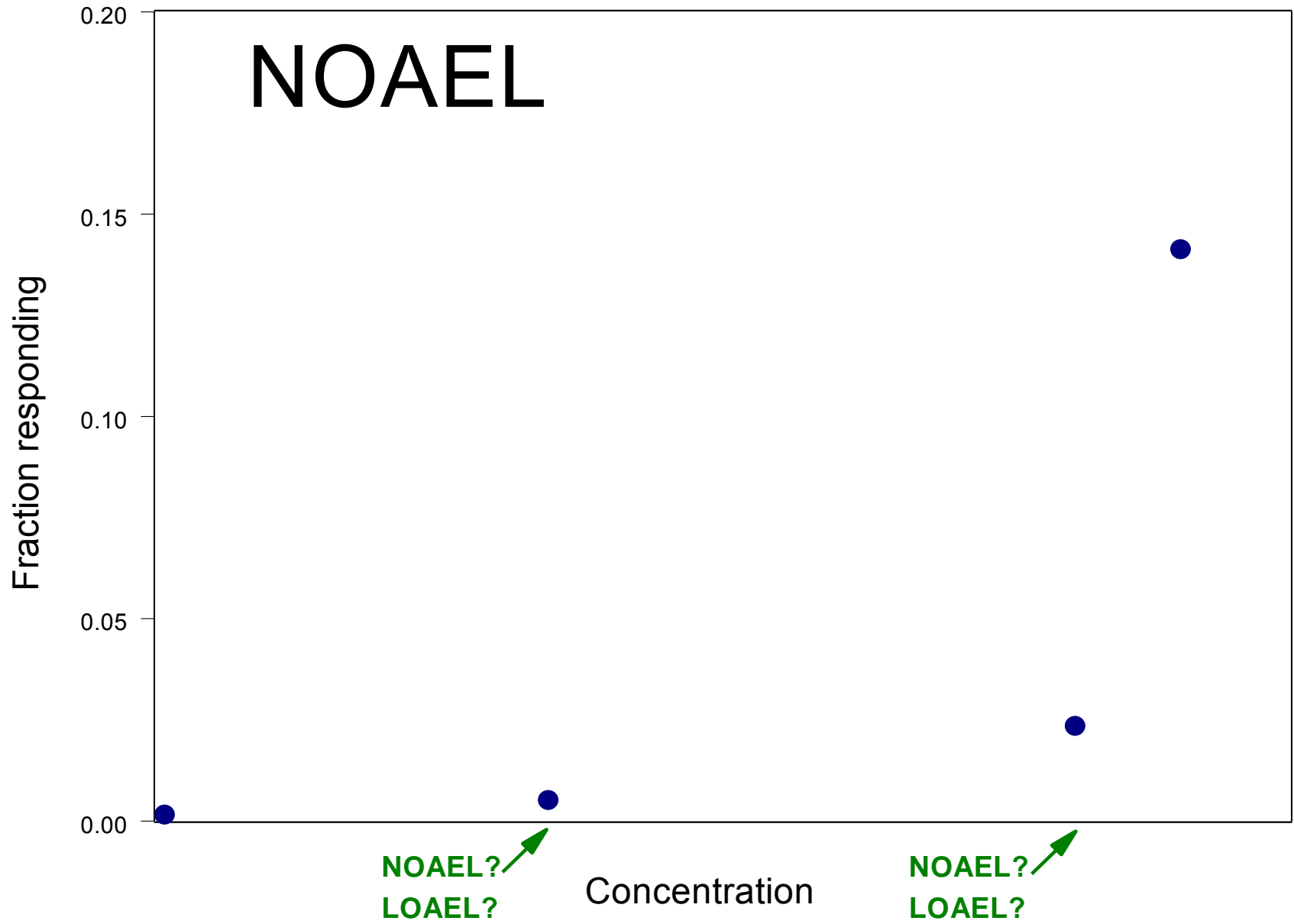
Chronological context

- **NOAEL (no observed adverse effect level)**
POD is chosen from administered doses (concentrations)
- **BMDS (benchmark dose software)**
POD is a dose (concentration) based on statistical modeling, for a single exposure duration
- **CatReg (categorical regression software)**
PODs for one or more user-specified exposure durations

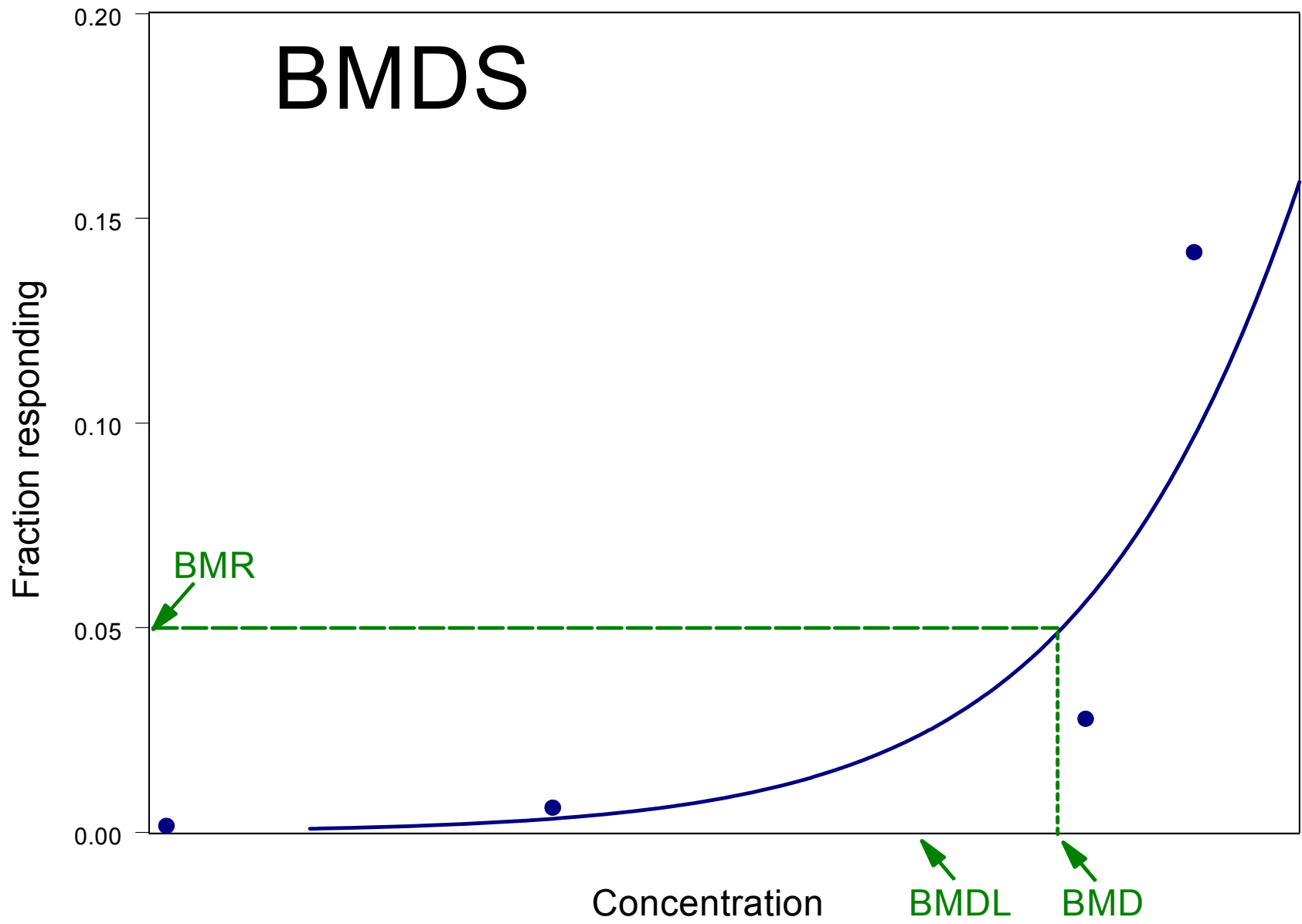
Definitions

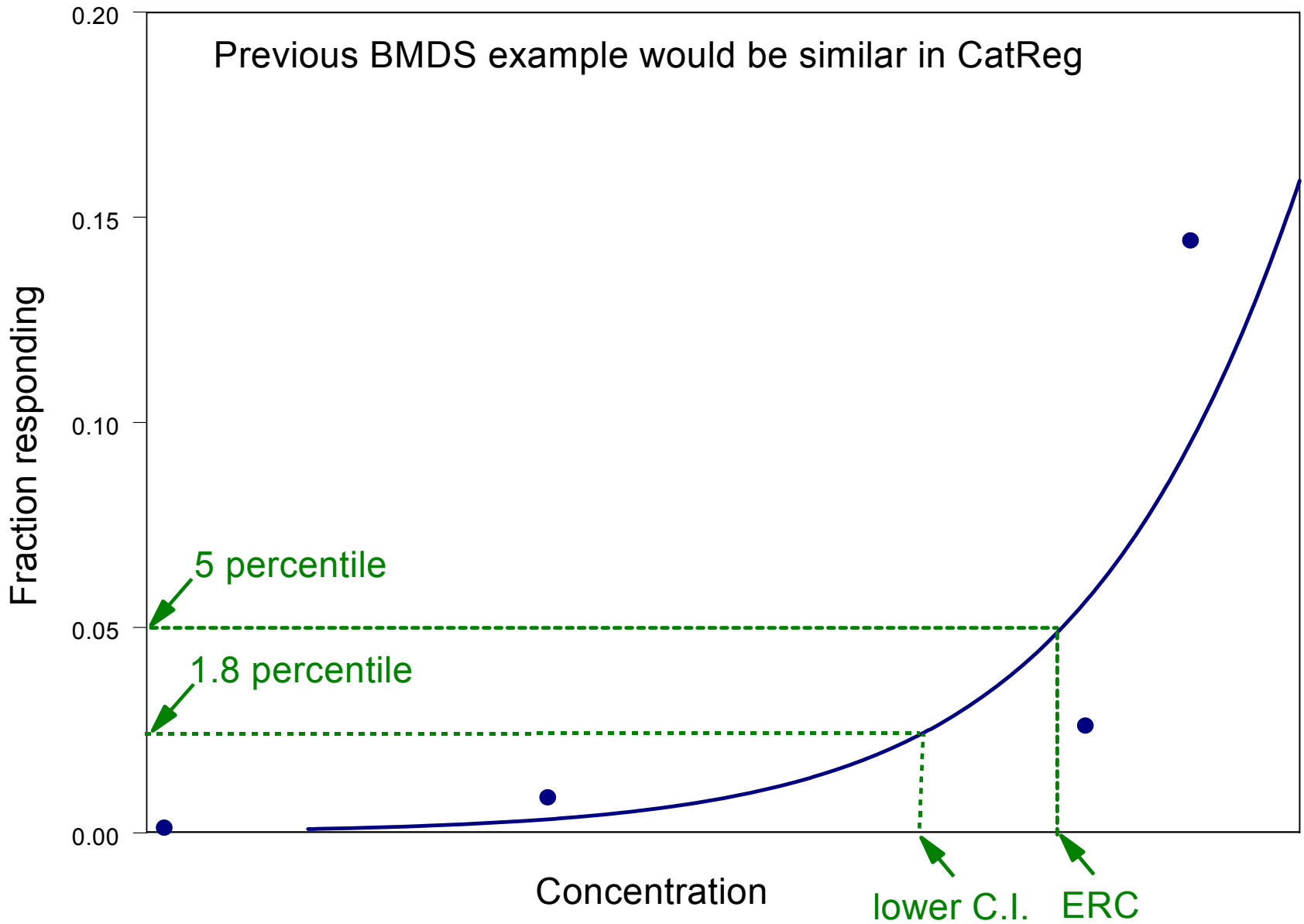
- **NOAEL, LOAEL:** The highest dose with no observed adverse effect level; lowest dose with observed adverse effect level.
- **Benchmark dose (BMD):** The dose estimate to increase response by a specified amount, called the **benchmark response (BMR)**, e.g., the BMD for BMR of 10% increase.
- **Benchmark dose lower bound (BMDL):** The statistical lower bound on the BMD, e.g. the 95% lower confidence bound.

NOAEL



BMDS





Outline

- Introduction/Overview
- Basic features of CatReg
- Using CatReg

CatReg Features

- Addresses severity response data (one or more response levels), e.g., 0 =no adverse effect, 1 = mildly adverse effect, 2 = highly adverse effect
- Addresses effect of exposure duration (**T** for time) as well as concentration level (**C** for concentration)
- Addresses other variables that might have an effect, e.g., gender, species, study
- May address multiple studies simultaneously (meta-analysis)

CatReg Features (cont'd)

- Tests hypotheses, e.g.,
 - H_0 : No gender effect
 - H_0 : Parallel response curves
- Estimates ERC (extra risk concentration) with C.I., e.g.,
at what concentration would 10% respond if exposed for 1hr?
- Estimates a C-T [also called C×T] relationship, e.g.,
if C is increased 25%, by how much does T need to be reduced to keep extra risk at the same level?

CatReg Limitations

- Availability of data may be limited and severity classification of effects may be difficult
- Analysis can be time-consuming
- Response data must be input as incidence of ordered categories. Continuous response data must be converted to categorical data. Generically reported response, e.g., “some immobility was observed”, is difficult to address quantitatively.

Outline

- Introduction/Overview
- Basic features of CatReg
- Using CatReg

Loading CatReg

(for Windows 98/2000/XP operating systems)

- (1) Download R from <http://cran.r-project.org/>
- (2) Create a sub-directory named “CatReg” like:
`C:\Program Files\R\r221\CatReg`. (Note: “r221” is the current R version)
- (3) Download attached source code files to “CatReg” directory: `catreg.R`, `ecdata.R`, `Util.R`, `Winplot.R`, and `Winstall.R`
- (4) Set the working directory in R to “CatReg”. Open R and enter command `setwd(“C:\\Program Files\\R\\r221\\CatReg”)`, then `getwd()` to verify that working directory is set
- (5) Select “File” from the main menu, and then “**Source R code...**” Go to the “`CatReg`” subdirectory, and double click “`Winstall.R`”. This will automatically source all the five CatReg source files
- (6) At the R prompt, type in “`catreg()`” to start CatReg

Overview of using CatReg to inform POD and C-T relationships

Assimilate data of interest

Sort data into subsets that are to be kept separate, as needed

Create an input file for each subset

Use CatReg to analyze data and generate parsimonious response model for each input file with sufficient data

Use CatReg to estimate extra risks with lower bounds, and the C-T relationships

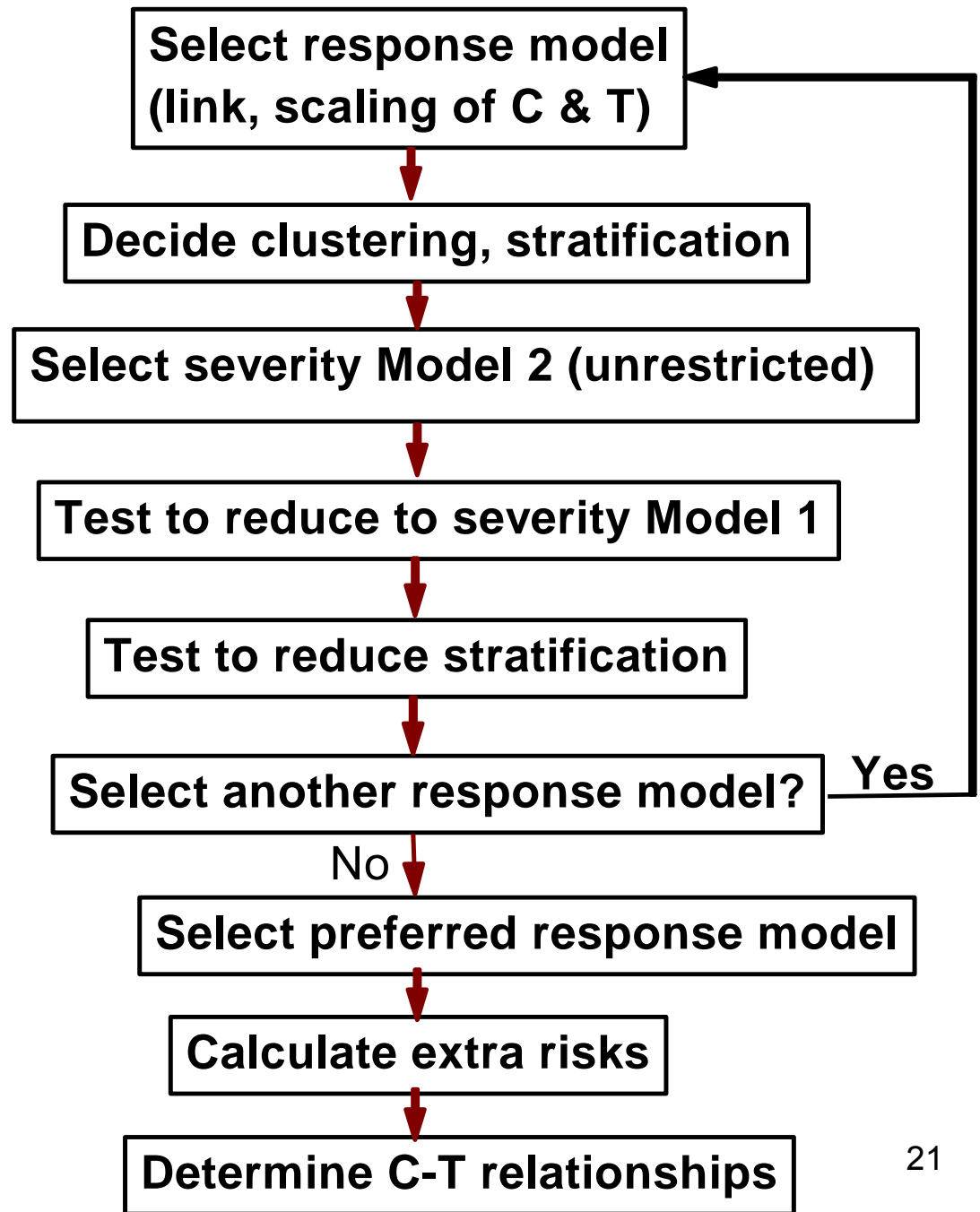
Use CatReg to estimate and test sources of differences between models, plot extra risk estimates and lower bounds

Summarize results to inform choice of a POD and C-T relationship

Entirely toxicologist
(not covered)

Expanded
on next
slide for
single input
file

Overview of analysis for a single input file



Specific objectives of this one-day class

- Introduce basic operational use of CatReg
- Address the steps described in the previous slide (overview of analysis for a single input file)
- Illustrate, with hands-on experience, features of CatReg with specially constructed input files
- Generate concentration-time (C-T) response curve(s)
- Calculate concentrations corresponding to specified extra risks and exposure durations (time, T), with confidence bounds
- Estimate the C-T relationship (e.g., “n” in C^nT)

Topics of a forthcoming advanced class (tentative)

- Generating data pool(s) for meta-analysis
 - testing to pool data (aside from quality)
 - detecting outlier studies and observations
- Estimating and testing of
 - intra-species variability
 - species differences
 - covariate effects
- Utilizing studies with limited data
- Testing model assumptions (probit, logit)
- Plotting to help assess model fit

This class: First step is to develop a parsimonious model for an input file

- A criteria for parsimony as used here:
AIC (Akaike Information Criteria) =
 $-2\log(\text{maximized likelihood function}) + 2x(\text{number of parameters used})$ [lower is better]
- Note that parameters are valued according to their contribution to maximizing the likelihood function (loosely, explaining response variability)
- A general approach: Put in parameters that might be important and reduce model by testing
- Compare results for different models by AIC (e.g., probit, logit)

Some precautions

- Don't violate biological plausibility
- Not rejecting an hypothesis does not mean accepting it
“Absence of evidence is not evidence of absence” C. Sagan?
- A series of simple tests is not a substitute for a joint test
- Axiom: There are never enough data! Consider objectives and extract as much information from the data as possible

Using CatReg

(general outline)

1. **Minimal setup** –
Just one adverse severity level (quantal response)
2. **Add adverse severity levels** –
Just intercept or all parameters differ by severity?
3. **Add time as a variable** –
How to model exposure as a function of both C and T?
4. **Example of categorizing continuous data** –
How to decide on break points?
5. **Add other variables that might have an effect** –
Which parameters are affected?

Using CatReg (cont'd)

(example motivated by propylene oxide)

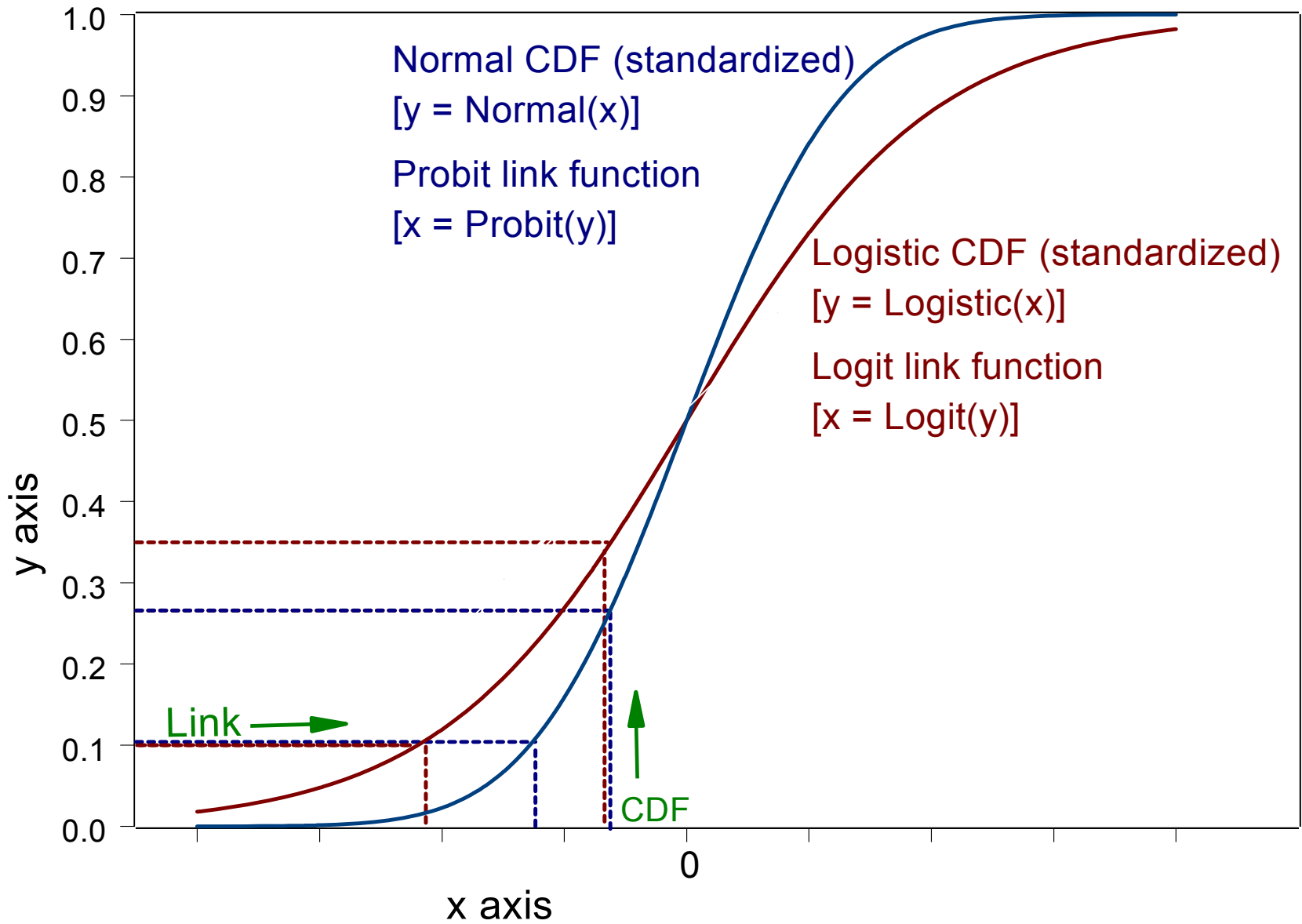
1. **Minimal setup** –
mice, 4hr exposure (alive, dead)
2. **Add adverse severity levels** –
same except (alive without dyspnea, alive with dyspnea, dead)
3. **Add time as a variable** –
same except 1hr, 2hr, 4hr, 8hr exposures
4. **NA** (applies to continuous data)
5. **Add other variables that might have an effect** –
same except mice and rats, by gender, are exposed

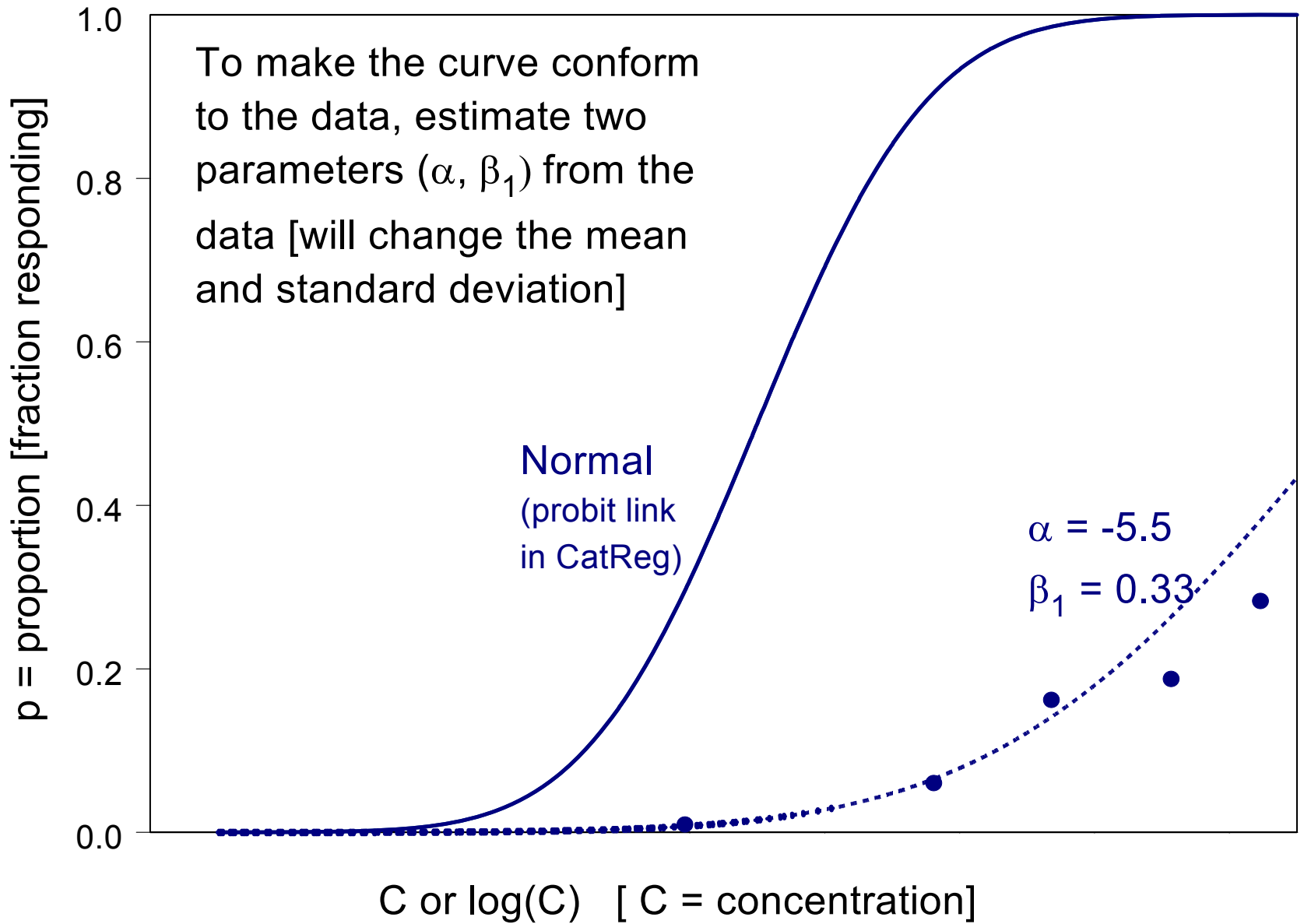
1. Minimal setup

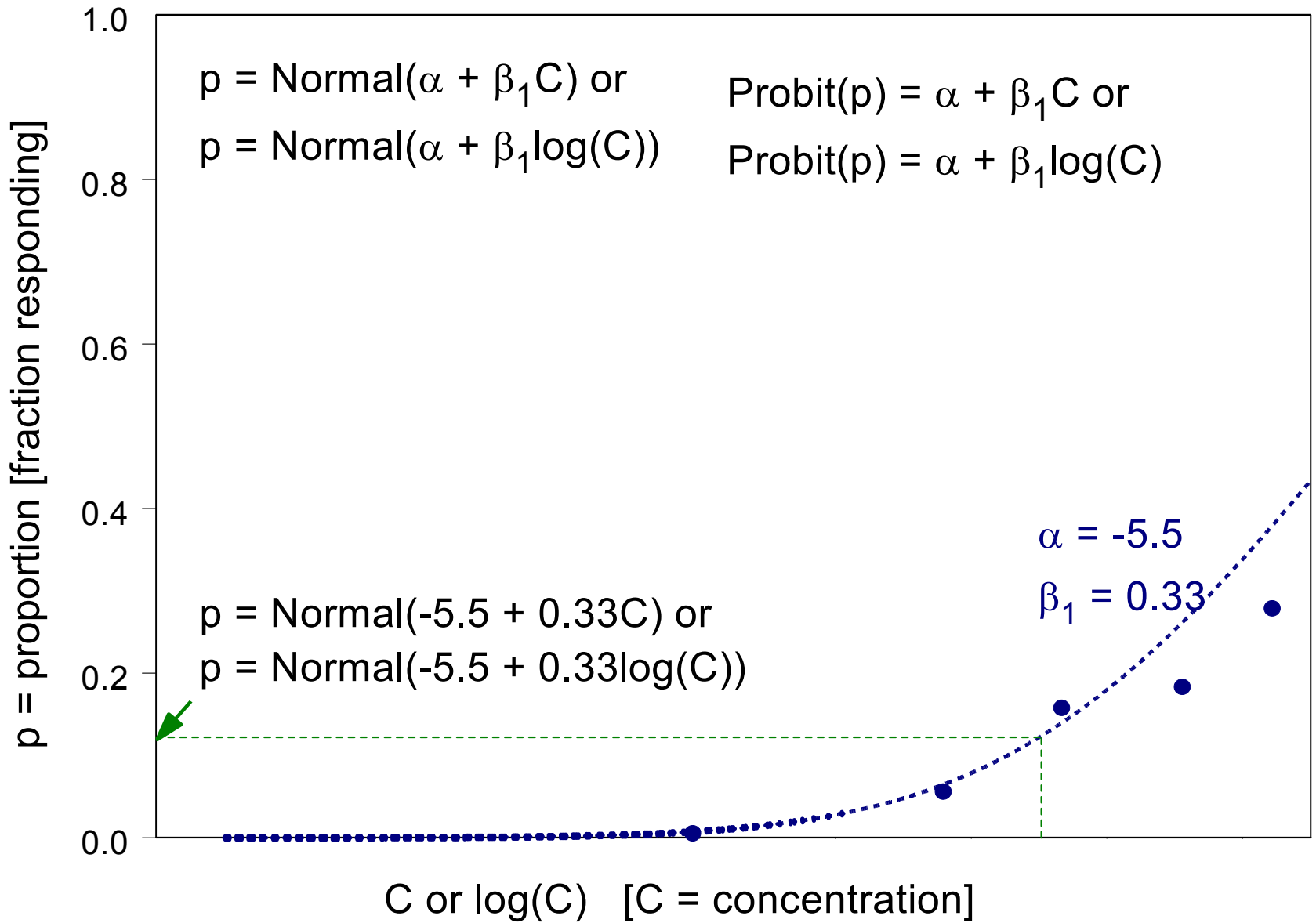
- **Description**
 - quantal response
 - no time variable (fixed exposure duration)
 - no other variables (species, etc.)
- **Two required options**
 - link function (probit or logit)
 - concentration scale (linear or \log_{10})
- **Correspondence to BMDS model options**
probit, logistic (same as logit here), log-probit, log-logistic

Basic concept

- Start with one of two basic functions (“models”) (cumulative distribution functions, mean = 0, scale = 1)
 - Normal function
 - Logistic function
- Their corresponding link functions are
 - Probit
 - Logit
- Statistically adjust parameters to fit response data







2. Add severity levels

- Let S be the number of adverse severity levels.
- Coding severity of response for CatReg might be:
 - $S=1$ (0= no adverse effect, 1= adverse effect)
 - $S= 2$ (0=none, 1=adverse, 2=lethal)
 - $S= 3$ (0=none, 1=mild/moderate, 2=severe, 3=lethal)
- Choose option for modeling relationship between severity levels (Model 1 or Model 2)

On severity >1

- **Two examples**
 - Propylene oxide, S=2 (0=negligible, 1=dyspnea, 2=lethal)
 - Potassium nitrate (in water), S=3, incidence of severity of adrenal lesions (0=none, 1= very slight, 2= slight,3=worse)
- **See User Manual** for a more detailed example
- **Precautions**
 - Separate endpoints, even with the same severity, should generally be identified separately in the input file (they are unlikely to have the same response curve, and CatReg can test if they do)
 - Response curves should not cross (e.g., probability of severity ≥ 2 should never be less than probability of severity ≥ 1). For example, some animals might die very soon from low exposure.

CatReg option when $S > 1$

- **Model 1: cumulative odds model.** Parallel curves parallel (shown for $S=2$)

$$p_1 = \text{Normal}(\alpha_1 + \beta_1 \log(C)) \text{ proportion } \geq \text{sev } 1$$

$$p_2 = \text{Normal}(\alpha_2 + \beta_1 \log(C)) \text{ proportion } \geq \text{sev } 2$$

(Normal or Logistic, $\log(C)$ or C)

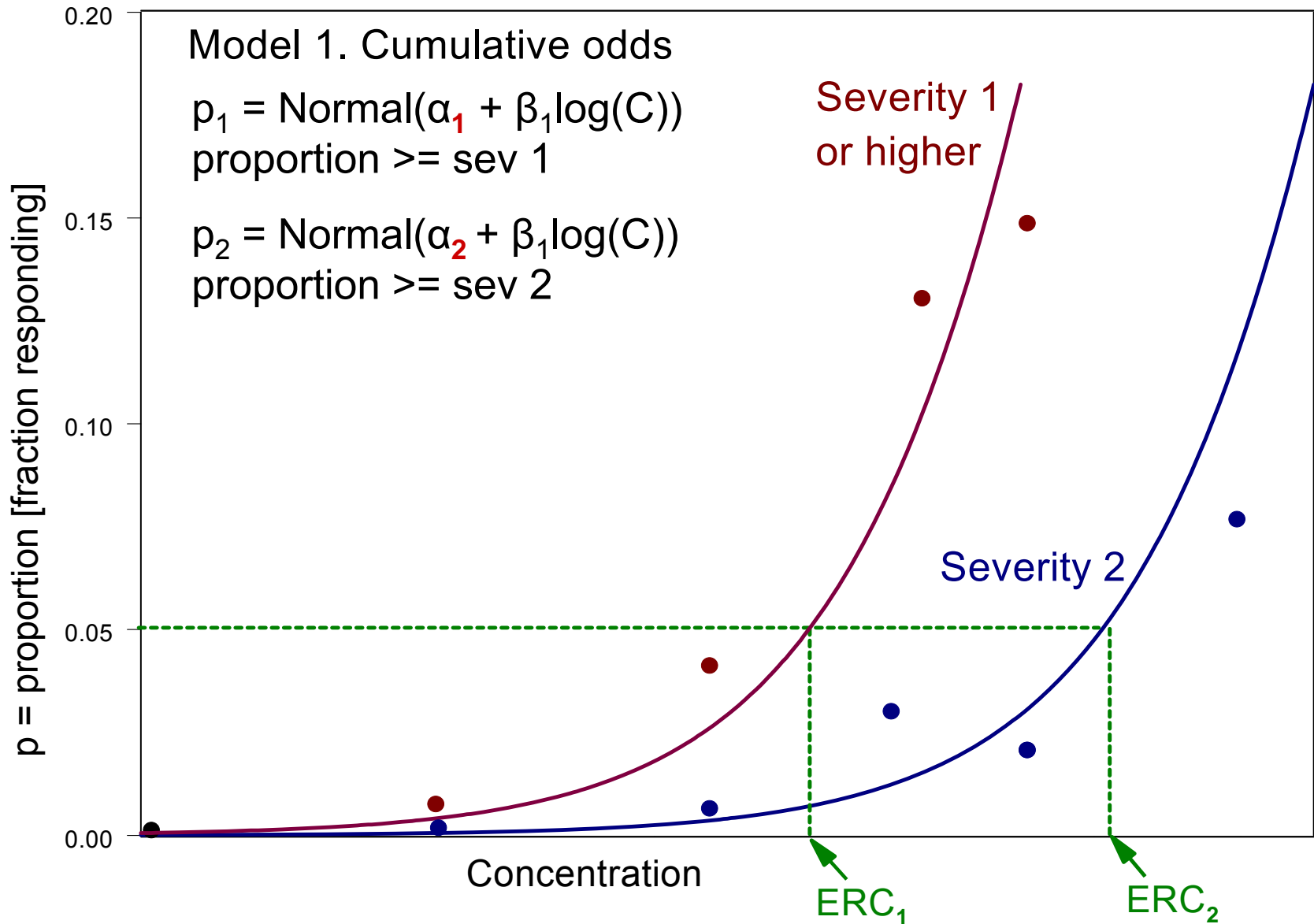
- **Model 2: unrestricted cumulative model.** Unrestricted curves ($S=2$)

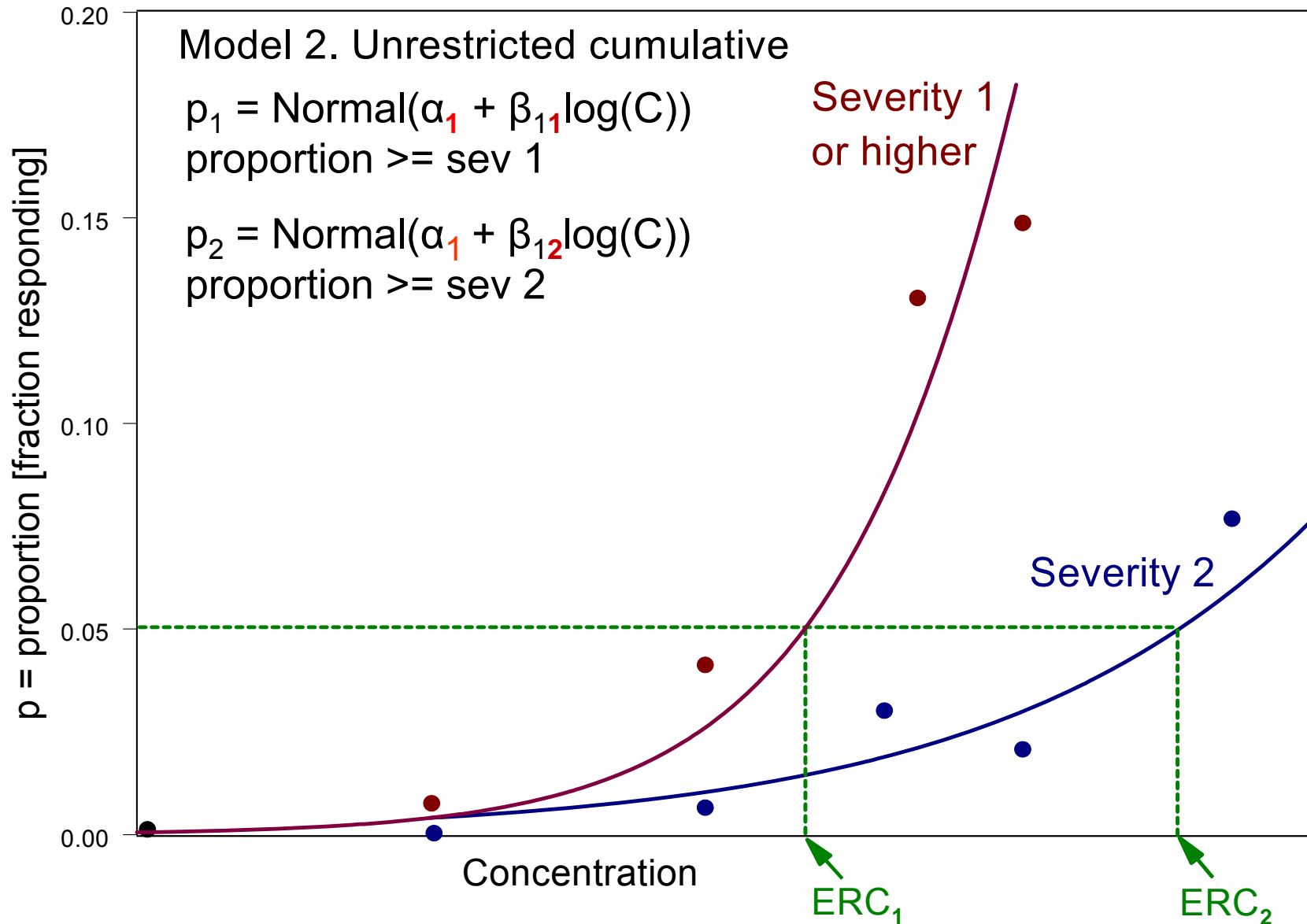
$$p_1 = \text{Normal}(\alpha_1 + \beta_{11} \log(C)) \text{ proportion } \geq \text{sev } 1$$

$$p_2 = \text{Normal}(\alpha_2 + \beta_{12} \log(C)) \text{ proportion } \geq \text{sev } 2$$

(Normal or Logistic, $\log(C)$ or C)

[Next two slides were not made with CatReg]





Next – we will work through an example with two adverse severity levels ($S = 2$)

First – some preparatory material



Input file format

(required variables)

- **Nsub** - number of subjects in treatment group
- **mg/m³** - concentration (C)
- **Hours** - time (T) (Not required if all times are the same)
- **SevLo** - severity (0,1, 2...)
(SevHi is added if the effect is given as a range, e.g. 1-2)
- **Incid** - incidence of severity

Input file example

First nine rows of file twosev.csv (coma delimited)

SevLo	Incid	Nsub	mg/m3	Group	<i>This column not part of input file</i>
0	10	10	1	1	mg/m3 - concentration (C)
0	7	10	2	2	Nsub - number of subjects
1	3	10	2	2	Incid - incidence
0	5	10	3	3	(must add up to Nsub)
1	5	10	3	3	SevLo - severity 0, 1, or 2
1	10	10	4	4	(SevHi is added when there
0	8	10	2	5	is a range, e.g., 1-2)
1	2	10	2	5	

Exercise

- Put file `twosev.csv` on your screen
If need name of working directory, go to R and then use
`>getwd()`
- Enter data for a 16th treatment group.
Number of subjects: 10
Exposure level: 16 mg/m³
Results: 1 subject at severity 1, 9 subjects at severity 2
- Save file in working directory

Input file example

Rows added to twosev.csv

SevLo	Incid	Nsub	mg/m3	Group	
1	1	10	16	16	
2	9	10	16	16	

Some notes on using CatReg

- “>” is the R prompt
- R is case-sensitive
- To call CatReg use R command
 >catreg() <CR>
- To get out of CatReg, use
 “Esc” key <CR>
- You can minimize CatReg and return to it later

Getting the input file to CatReg

- For this class, it will be easiest if your Excel input files are in your working directory and you put a shortcut to that directory on the windows display
- To put the shortcut on the windows display
 - > `getwd()` [R command (not in CatReg) will get name of working directory if needed]
 - Go to the working directory (e.g., use windows explorer) and right click on it. Select “create shortcut”. Drag the shortcut to windows display
- The input file for Problem 1 is `twosev.csv`, an Excel file
- At the Catreg prompt, enter the input file name
>Name of data file? `twosev.csv`

Getting the input file to CatReg (cont'd)

Some options if the input file is not in the working directory, e.g., is in C:\CRTProb2\twosev.csv.

- Key in the entire path “C:\CRTProb2\twosev.csv”
- Create a shortcut on the windows display for C:\CRTProb2 (in which case you can right-click, left-click on properties, copy and paste “C:\CRTProb2” into CatReg and append “twosev.csv”
- Create a shortcut on the windows display for the input file itself “timesev.csv” (in which case you can right-click, left-click on properties, copy and paste “C:\CRTProb2\twosev.csv”

Steps during CatReg execution

1. CatReg - reads input file
2. User - sets options
3. CatReg - estimates response curve, with output to monitor and/or output file (in working directory)
4. User - settings for extra risk* (values go to output file)
5. User - initiates commands of interest (plots, tests, other) (goes to monitor)

*We will assume that an adverse response is entirely due to the administered chemical, i.e. no “background” risk. Then extra risk is just the proportion (fraction) responding.

Problem 1

(objective)

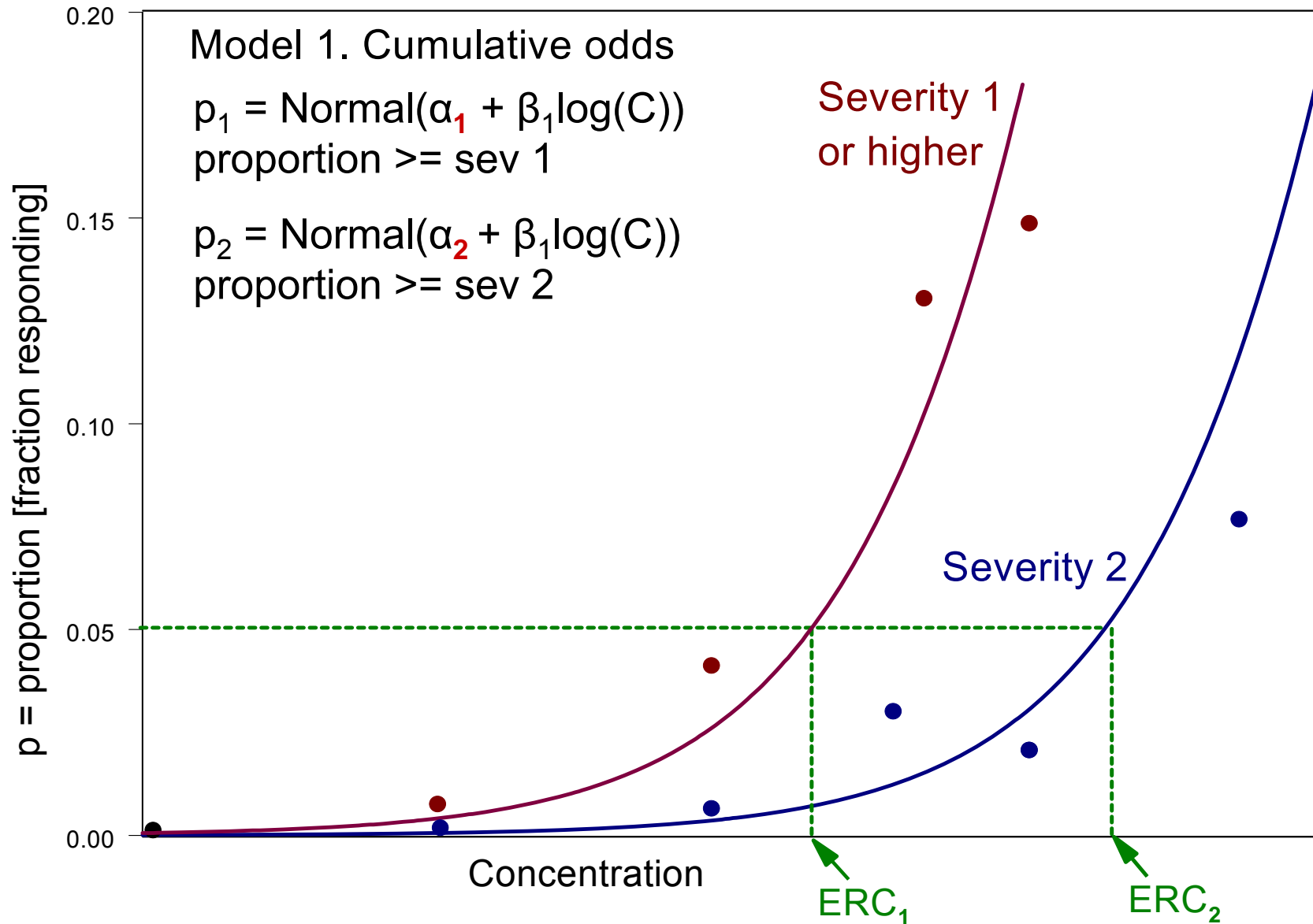
- Data set: `twosev.csv` (if you did not get the exercise correct, use `twosevcorr.csv`, which is correct)
- Use $\log(C)$ and logit link (logistic) [like log-logistic model in BMDS, except we have two adverse severity levels]
- Fit Model 2 (unrestricted cumulative) and test whether it can be reduced to Model 1 (cumulative odds)
- Estimate ERC10, the concentration at which extra risk is 0.10, and its 95% one-sided lower bound, for both severity levels

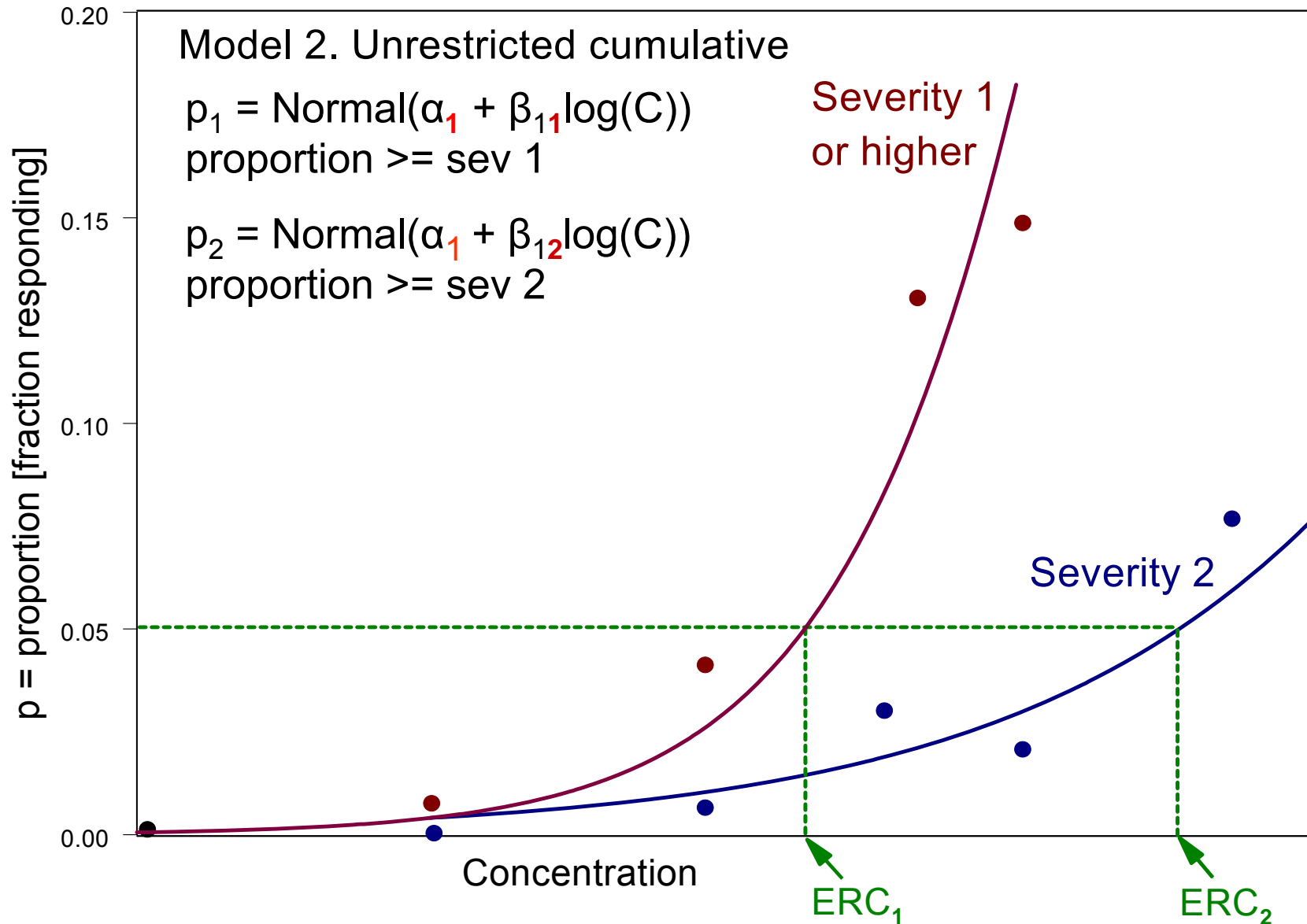
Recall Models 1 and 2 and example plots

(to be shown again in the next two slides)

- **Model 1: cumulative odds model.** Parallel curves
 $p_1 = \text{Logistic}(\alpha_1 + \beta_1 \log(C))$ proportion \geq sev 1
 $p_2 = \text{Logistic}(\alpha_2 + \beta_1 \log(C))$ proportion \geq sev 2
- **Model 2: unrestricted cumulative model.** Unrestricted curves
 $p_1 = \text{Logistic}(\alpha_1 + \beta_{11} \log(C))$ proportion \geq sev 1
 $p_2 = \text{Logistic}(\alpha_2 + \beta_{12} \log(C))$ proportion \geq sev 2

[Next two slides were not made with CatReg]





Problem 1

(running CatReg)

- The monitor display for running Problem 1 will be shown in the next few slides
- **This color is for user**, this color for CatReg, **this color for inserted commentary**
- Please launch R on your computer
- Key in the instructions on your notebook as we go

> catreg() <CR>

Name of data file? `twosev.csv` <CR>

[Next - option to filter, i.e., ignore some data, e.g. eliminate a study, outliers]

Variables: SevLo Incid Nsub mg/m3 Incid GpSize

Filter the data on which variables? (none):

1: <CR>

[Next - option to cluster, i.e., account for variability between samples, e.g. study]

Variables: SevLo Incid Nsub mg/m3 Incid GpSize

Define clusters using which variable(s)? (none):

1: <CR>

[Next - option to stratify, i.e., estimate variable effects, e.g., species, gender]

Variables: SevLo Incid Nsub mg/m3 Incid GpSize

Stratify intercepts on which variables? (none):

1: <CR>

Variables: SevLo Incid Nsub mg/m3 Incid GpSize

Stratify concentration (mg/m3) on which variables? (none):

1: <CR>

[Next - option to log-transform concentration]

Log or linear scale for mg/m3?

1: mg/m3

2: log10(mg/m3)

Selection: `2` <CR>

[Next - choose probit (normal) or logit (logistic) response model]

Link function?

1: logit

2: probit

3: cloglog

Selection: 1 <CR>

[Next - choose to make response curves for different severity levels parallel (Model 1) or unrestricted (Model 2)]

Model?

1: cumulative odds model

2: unrestricted cumulative model

Selection: 2 <CR>

[Next - option for background risk. If “no”, a background concentration is estimated. Option only appears when log(C) is used. Is automatic otherwise]

Assume zero background risk

(i.e. response cannot occur at zero concentration)? (y): y <CR>

[Next - name the output file or use default]

Name of the output? (twosev): <CR>

Output file is twosev.out

Continue? (y): <CR>

[Next - option to write covariance matrix to output file]

Printout covariance matrix? (n): <CR>

[Next - some summary info]

Input file : twosev.csv

Filtered data: none

Model : unrestricted cumulative model

Link : logit

Clustering : none

Message :

Iterations : 17 8

Deviance : 199.4955

Residual DF : 28

AIC : 207.4955 [Akaike Information Criteria, used to compare how well data fit for different option settings, but always with the same data set. Lower is better]

Scale:

Concentration: log10(mg/m³)

Duration : Hours

Stratification:

No Stratification on Intercept, Concentration and Duration.

[Next - each parameter estimate, its standard error, Z value for testing the hypothesis the parameter is zero, and p-value of the test. The formula for the model is inserted here for easy reference; it does not appear on your screen]

[The following Model 2 description is an insert for reference]

Model 2: unrestricted cumulative model. Unrestricted curves

$$p1 = \text{Logistic}(\alpha_1 + \beta_{11}\log(C)) \text{ proportion } \geq \text{sev } 1$$

$$p2 = \text{Logistic}(\alpha_2 + \beta_{12}\log(C)) \text{ proportion } \geq \text{sev } 2$$

Coefficients:

		Est.	S.E.	Z-Test =0	p-value
α_1	SEV1	-3.56	0.79	-4.46	1e-05
α_2	SEV2	-7.31	1.19	-6.12	1e-05
β_{11}	LG10CONC:SEV1	8.71	1.55	5.59	1e-05
β_{12}	LG10CONC:SEV2	7.67	1.30	5.90	1e-05

[Next - analysis of deviance table]

Analysis of Deviance Statistics:

Generalized R-squared: 0.4

	DF	Deviance	Mean.Dev	Gen.F	pvalue
Model	2	132.7647	66.382	9.317	8e-04
Residual	28	199.4955	7.125		
Total	30	332.2602			

#####Note: About 40 % of the variation in the response is accounted for by the explanatory variables in the current model fit. The p-value of the model fit is ≤ 0.05 . This is generally considered significant, indicating that the current model fit is acceptable.

#####

[Next - set options to calculate ERC for severity >=1.]

Percentile for ERC? (default=10): <CR>

Severity level for ERC? (default = 1): <CR>

Percentile for 1-sided upper and lower confidence intervals? (default=95): <CR>

[Next - option to write 20 ERC values

(for range of time divided into 19 intervals) to a text file]

Write ERC data to text files? (n):n <CR>

[Next - set options to calculate ERC for severity >=2.

Use `ecdata()` command whenever need to reset ERC settings]

> `ecdata()`<CR>

Percentile for ERC? (default=10): <CR>

Severity level for ERC? (default = 1): 2 <CR>

Percentile for 1-sided upper and lower confidence intervals? (default=95):

Write ERC data to text files? (n): <CR>

[Note - For the response to the query from CatReg we opted for severity 1.

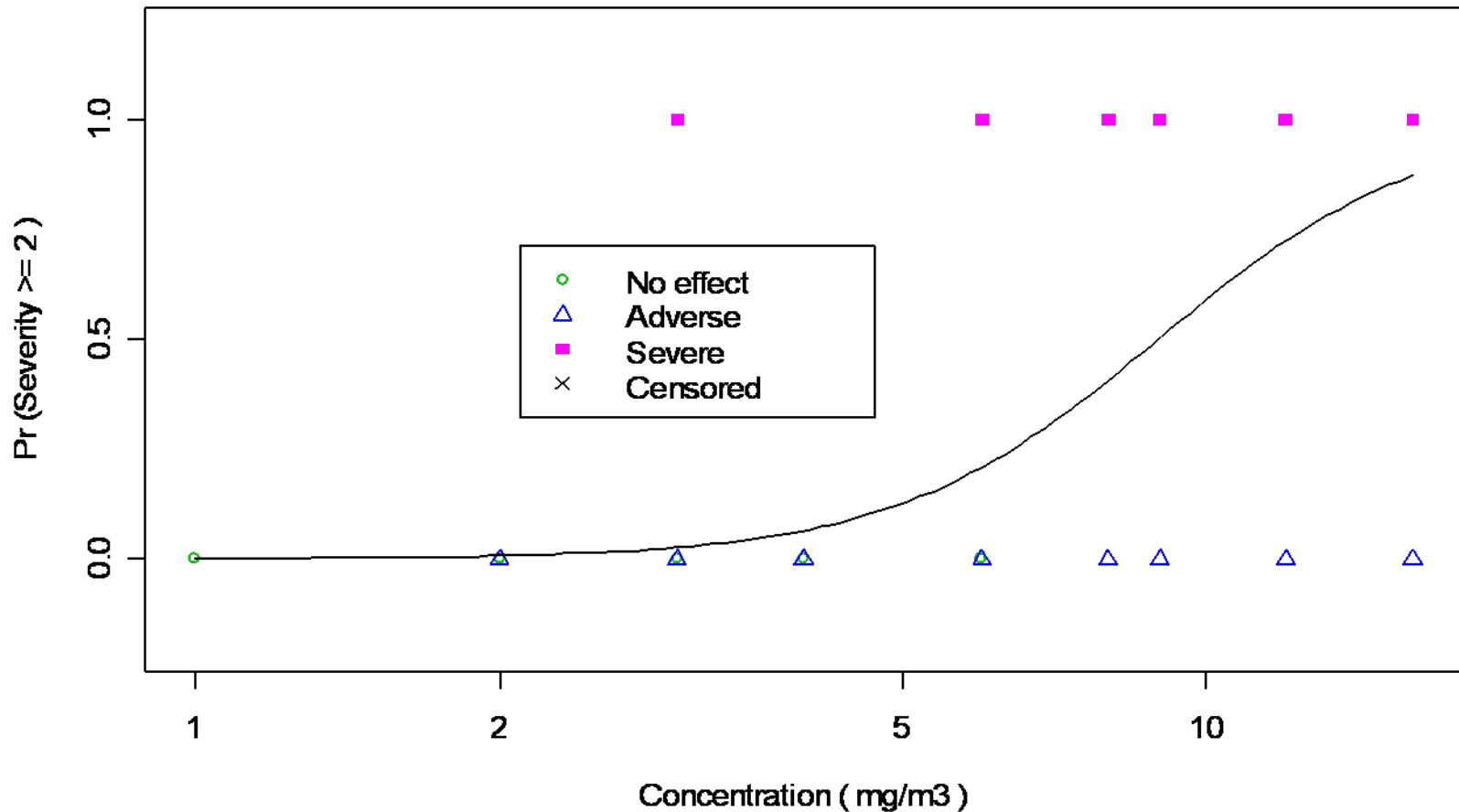
Those ERC values are written to the output file for times 1,4,8,24 hours

(although time does not enter this problem). To include severity 2 as well, we used `ecdata()` and opted for severity 2. Whenever you use `ecdata()`, the ERC values are added to the output file for times 1, 4, 8, 24 hours]

>prplot(time=) <CR> [Sev2 appears because it is the current setting. To change severity level, use ecdata()]

Duration (Hours) = 0 Stratum = SEV2

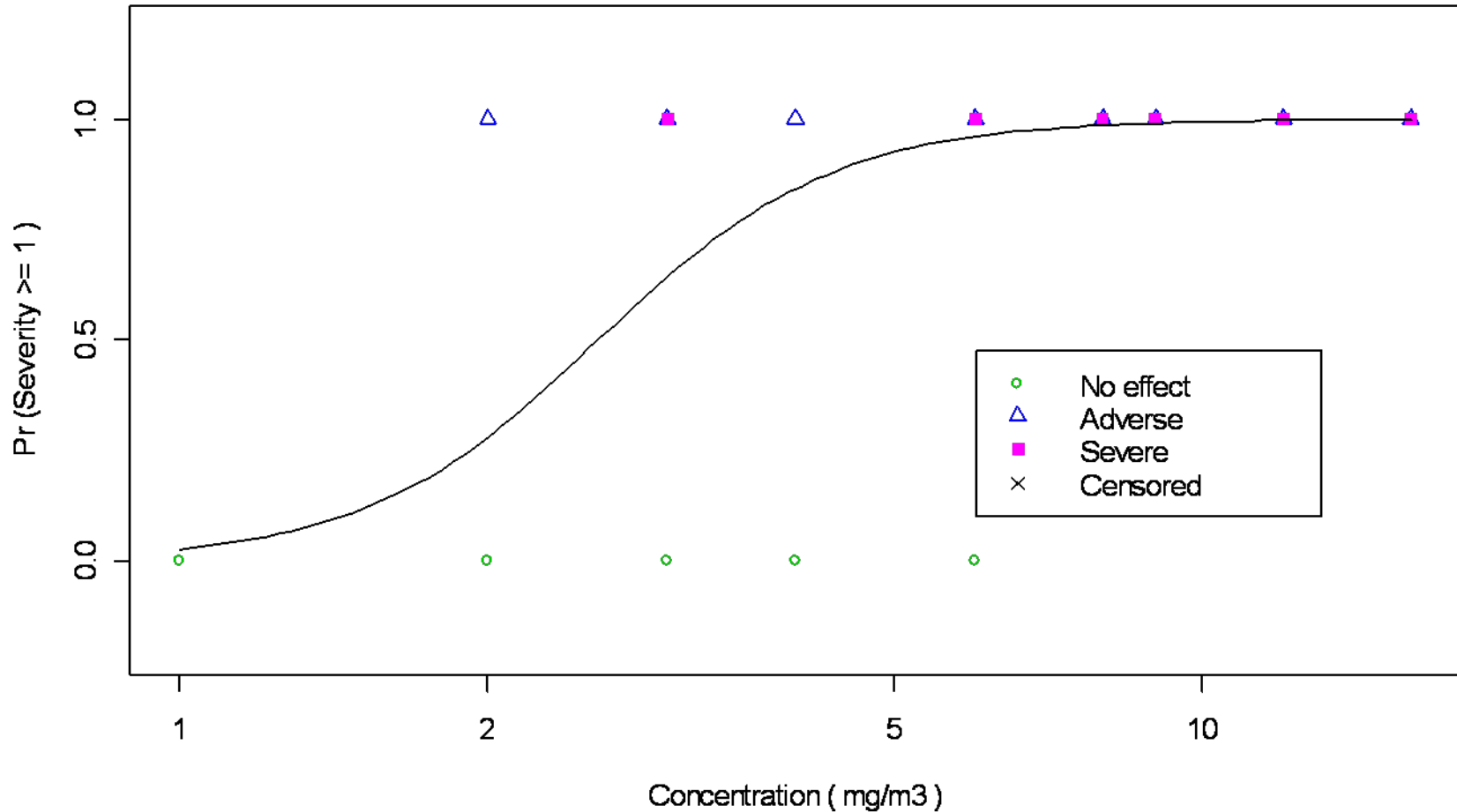
: When both concentration and time are not specified, Time=1(when log(Time)), Time=0(when Time) for Probability



>prplot(time=) <CR> [First used ecdata() to change the severity level to Sev1.
That caused the ERC values for Sev1 to be written to the output file again]

Duration (Hours) = 0 Stratum = SEV1

: When both concentration and time are not specified, Time=1(when log(Time)), Time=0(when Time) for Probability



[Next - test hypothesis that Model 1 is adequate, i.e., that we can change from Model 2 to Model 1. See the next two slides for review of graphs for Models 1 and 2. In terms of parameters, the hypothesis being tested is $H_0: \beta_{11} = \beta_{12}$]

> `parallel.test()`

grps: 1

Hypothesis Test Results

Coefficients tested for removal: None

Coefficients tested for equality:

Group 1 : [β_{11}] LG10CONC:SEV1 [β_{12}] LG10CONC:SEV2

Test statistics:

Chisquare	df	p-value
0.2598078	1	0.61025

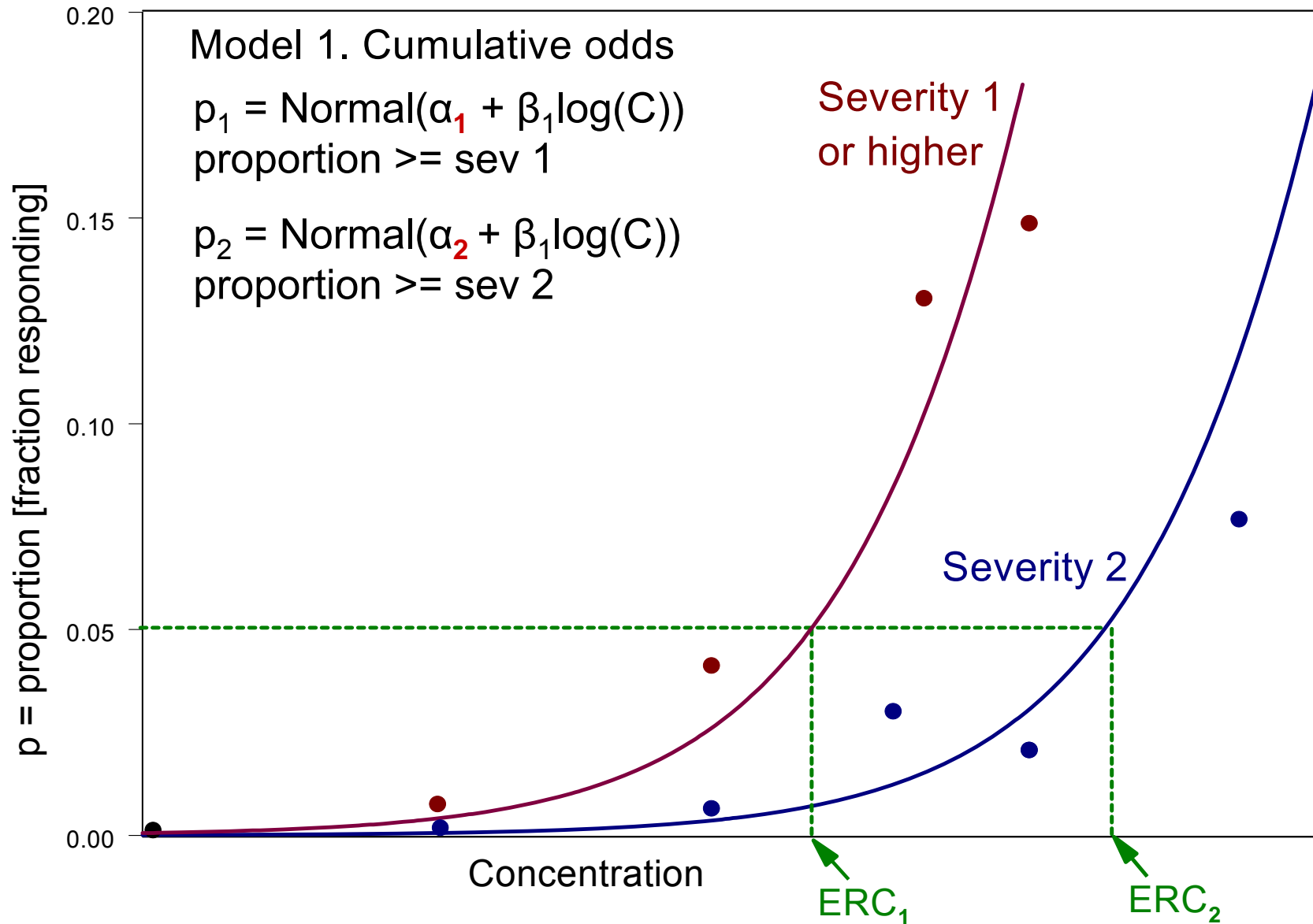
#####

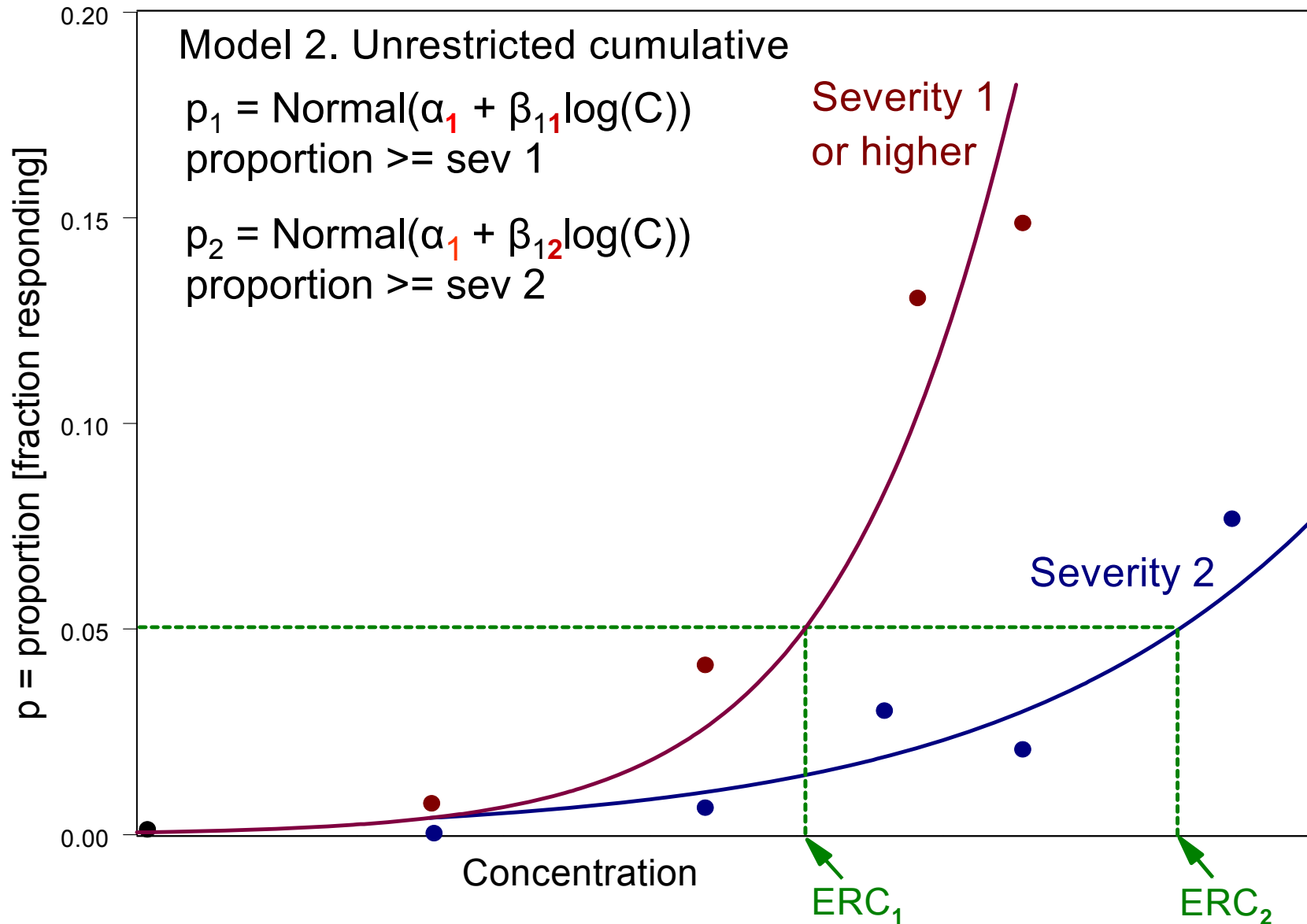
The P Value of the equality test is > 0.05. This is generally considered not significant, indicating that it would be more appropriate to use the simpler restricted cumulative model.

(See Section 5.2 of "CatReg Software Documentation")

#####

[Next two slides were not made with CatReg]





Problem 1

(access output file for Prob 1)

- Please open working directory (left click on shortcut on windows display)
- Look for file with output name (default name was twosev.out) and open it (left click)
- For this problem, it mostly repeats what was shown on monitor
- The ERC values for selected times, 1, 4, 8, 24 hours are always printed in output file (time is not in this problem)

Problem 1

(ERC summaries printed to output file)

[Values from ERC settings (severity 1, extra risk 0.1, 95% upper and lower bounds) in response to CatReg query. Times shown are automatic]

***** ERC summary *****

< ERC10 at specific time point(s) >

: One-sided 95% lower bound and one-sided 95% upper bound [default values]

* Risk Type : extra risk [the only risk type currently available in CatReg]

* ERC10 at SEV1

	ERC10 Lower Bound of ERC10	Upper Bound of ERC10	
1 Hours	1.4346	1.1176	1.8414
4 Hours	1.4346	1.1176	1.8414
8 Hours	1.4346	1.1176	1.8414
24 Hours	1.4346	1.1176	1.8414

Note: These One-sided 95% lower bound and one-sided 95% upper bound confidence intervals are equivalent to the lower bound and upper bound of two-sided 90% confidence intervals.

Problem 1

(ERC summaries printed to output file - cont'd)

[Values from using ecddata() to alter ERC settings]

***** ERC summary *****

< ERC10 at specific time point(s) >

: One-sided 95% lower bound and one-sided 95% upper bound

* Risk Type : extra risk

* ERC10 at SEV2

ERC10 Lower Bound of ERC10 Upper Bound of ERC10

1 Hours	4.6375	3.8369	5.6051
4 Hours	4.6375	3.8369	5.6051
8 Hours	4.6375	3.8369	5.6051
24 Hours	4.6375	3.8369	5.6051

Note: These One-sided 95% lower bound and one-sided 95% upper bound confidence intervals are equivalent to the lower bound and upper bound of two-sided 90% confidence intervals.

Some comparisons if Model 2 is replaced by Model 1

Recall: `parallel.test()` indicated no significant difference between Model 1 and Model 2

	<u>Model 2</u>	<u>Model 1</u>
AIC	207.5	205.8
R ²	0.40	0.40
ERC10: Sev1	1.44 (1.12, 1.84)	1.36 (1.11, 1.68)
ERC10: Sev2	4.64 (3.84, 5.61))	4.79 (4.13, 5.56)

Exercise 1

Use probit link instead of logit and compare

- Repeat CatReg run of Problem 1 (input file twosev.csv) using probit link, $\log(C)$, Model 1
- Calculate ERC10, 95% one-sided lower confidence bound, sev 1 and sev 2

Some comparisons if logit link is replaced by probit link, for Model 1

	Probit <u>Model 1</u>	Logit <u>Model 1</u>
AIC	206.9	205.8
R ²	0.395	0.399
ERC10: Sev1	1.32 (1.07, 1.63)	1.36 (1.11, 1.68)
ERC10: Sev2	4.62 (4.01, 5.32)	4.79 (4.13, 5.56)

Very little difference!

Note on how CatReg saves to the output file

- CatReg will write over an output file of the same name, so use different output file names for those you want to save
- When you use the command `ecdata()` to add a new set of ERC values and confidence bounds, the results will be added to the existing output file

3. Add time as a variable ($S \geq 1$)

- Example of how one might proceed
- A C-T relationship is any function of (C,T) such that values of (C,T) for which it is constant produce the same risk (proportion responding), e.g.,

$$C^n T,$$

where “n” is an unknown parameter (suggested by ten Berge et al., 1986; Haber, 1924, with $n = 1$)

- Replacing C with $C^n T$ in $\alpha + \beta_1 \log(C)$ of the previous example, and redefining parameters, gives

$$\alpha + \beta_1 \log(C) + \beta_2 \log(T)$$

(β_1 is different, α is not) with $n = \beta_1/\beta_2$

CatReg options

- Four choices of “C-T metrics”

$$\alpha + \beta_1 \log C + \beta_2 \log T$$

$$\alpha + \beta_1 C + \beta_2 \log T$$

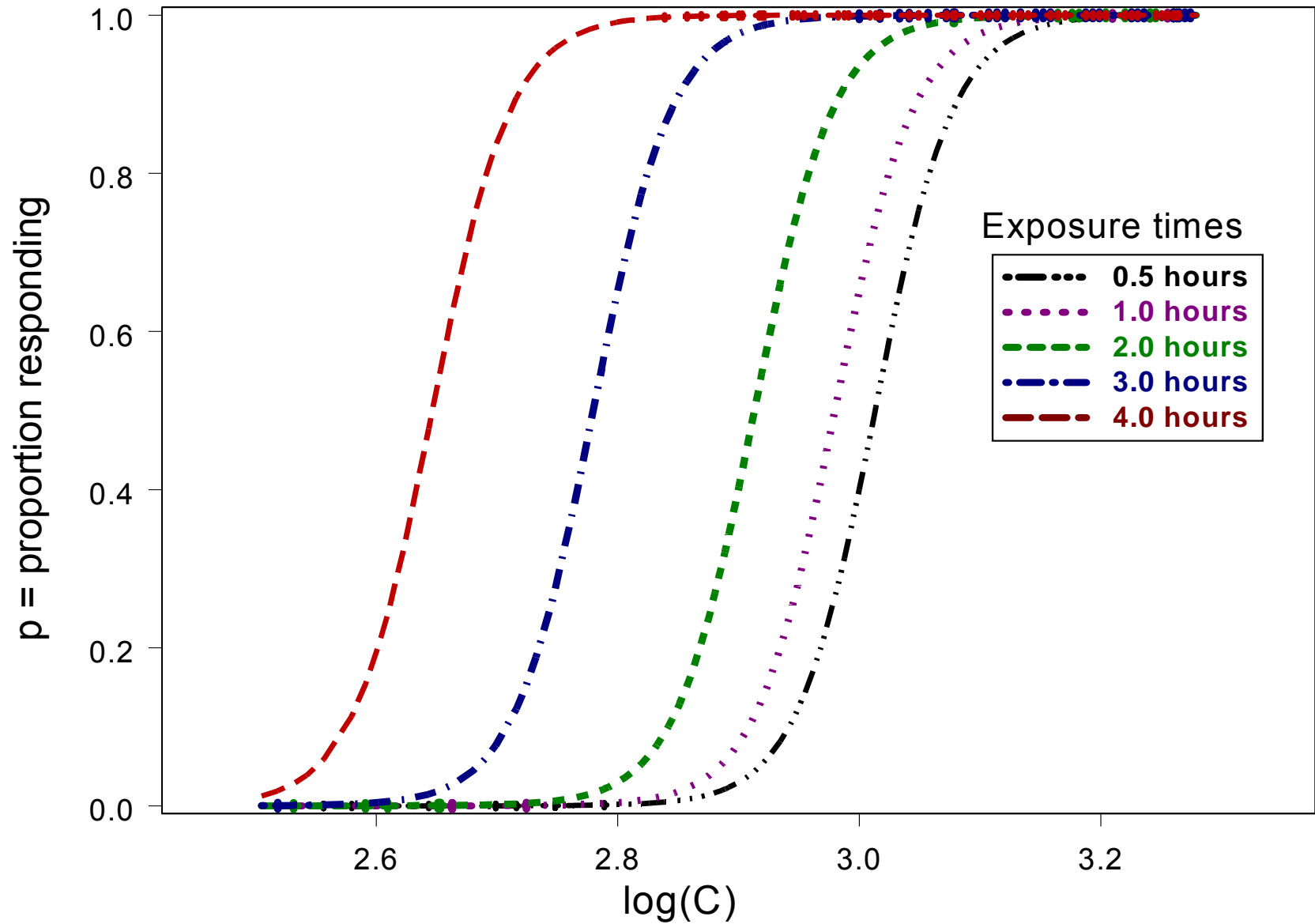
$$\alpha + \beta_1 C + \beta_2 T$$

$$\alpha + \beta_1 \log C + \beta_2 T$$

- Choice is determined by whether user chooses linear or log scale for C and/or T
- Recommended as a default

$$\alpha + \beta_1 \log C + \beta_2 \log T$$

[Next slide was not made with CatReg]



Exercise 2

Complete an input file from study data

- This is a hypothetical case for which you are asked to complete an input file
- A study was conducted to assess the toxicity of a chemical on rats
- Sixteen groups of 10 rats each were exposed by inhalation to concentrations of 1.0, 2.0, 3.0, 4.0 mg/m³ (25° C), at exposure durations of 1.0, 2.0, 3.0, 4.0 hours (one group per concentration/duration combination)
- The test rats were then examined for adverse effects on the lungs and each rat was classified as having “no effect”, “small effect”, or “large effect”

Exercise 2

Study data

			No	Small	Large					No	Small	Large
Group	mg/m ³	Hour	effect	effect	effect		Group	mg/m ³	Hour	effect	effect	effect
A	1	1	10	0	0		C	3	1	1	7	2
	1	2	7	3	0			3	2	1	7	2
	1	3	5	5	0			3	3	0	4	6
	1	4	0	10	0			3	4	0	5	5
B	2	1	8	2	0		D	4	1	0	10	0
	2	2	3	7	0			4	2	0	3	7
	2	3	2	7	1			4	3	0	2	8
	2	4	0	7	3			4	4	0	1	9

Exercise 2

Complete the file

- Put file exercise2.csv on your screen. (Minimize R, go to working directory, click on exercise2.csv)
- Add the data for Group D, 4 hours exposure, from the study report for Group D (shown below)

Group	mg/m3	Hour	No effect	Small effect	Large effect
D	4	1	0	10	0
	4	2	0	3	7
	4	3	0	2	8
	4	4	0	1	9

Exercise 2

Completed file

Group	mg/m3	Hours	SevLo	Incid	Nsub
A	1	1	0	10	10
		•			
		•			
		•			
D	4	1	1	10	10
D	4	2	1	3	10
D	4	2	2	7	10
D	4	3	1	2	10
D	4	3	2	8	10
D	4	4	1	1	10
D	4	4	2	9	10

Problem 2

(objectives)

- Input file: `sevtime.csv`
- Use $\log(C)$, $\log(T)$ and probit link in CatReg
- Interpret the summary table of estimates
- Estimate the concentration (ERC) at which extra risk is 0.10, with 95% lower bound, for selected times, by severity level.
- Discuss use of the C-T relationship
- Make several plots

Problem 2

(monitor display)

Name of data file? `severtime.csv`

>Log or linear scale for Hours?

1: Hours

2: $\log_{10}(\text{Hours})$

Selection: `2 <CR>`

>Model?

1: cumulative odds model

2: unrestricted cumulative model

Selection: `2 <CR>`

[Next - ERC summaries in output file]

* ERC10 at SEV1

	ERC10	Lower Bound of ERC10	Upper Bound of ERC10
1 Hours	1.3464	1.0527	1.7220
4 Hours	0.4239	0.3050	0.5890
8 Hours	0.2378	0.1515	0.3734
24 Hours	0.0952	0.0486	0.1864

* ERC10 at SEV2

	ERC10	Lower Bound of ERC10	Upper Bound of ERC10
1 Hours	3.477	2.902	4.167
4 Hours	1.611	1.309	1.982
8 Hours	1.096	0.785	1.530
24 Hours	0.595	0.343	1.034

Note:

1. These One-sided 95% lower bound and one-sided 95% upper bound confidence intervals are equivalent to the lower bound and upper bound of two-sided 90% confidence intervals.

“n” in $C^nT = k$

[New - Not in your version of CatReg]

* n at SEV1

n	Lower Bound of n	Upper Bound of n
1.18	0.88	1.47
	90% C. I. (0.88, 1.47)	95% C.I. (0.83, 1.52)

* n at SEV2

n	Lower Bound of n	Upper Bound of n
1.79	1.10	2.47
	90% C. I. (1.10, 2.47)	95% C.I. (0.97, 2.61)

Note:

These one-sided 95% lower bound and one-sided 95% upper bound confidence intervals are equivalent to the lower bound and upper bound of two-sided 90% confidence intervals.

[Confidence bounds on “n” do not provide confidence bounds on extra risk]

[Next – return to CatReg to request ERC10 values at some specific durations]

```
> ec3table(duration = c(1.0,1.25,1.5,1.75,2.0))
```

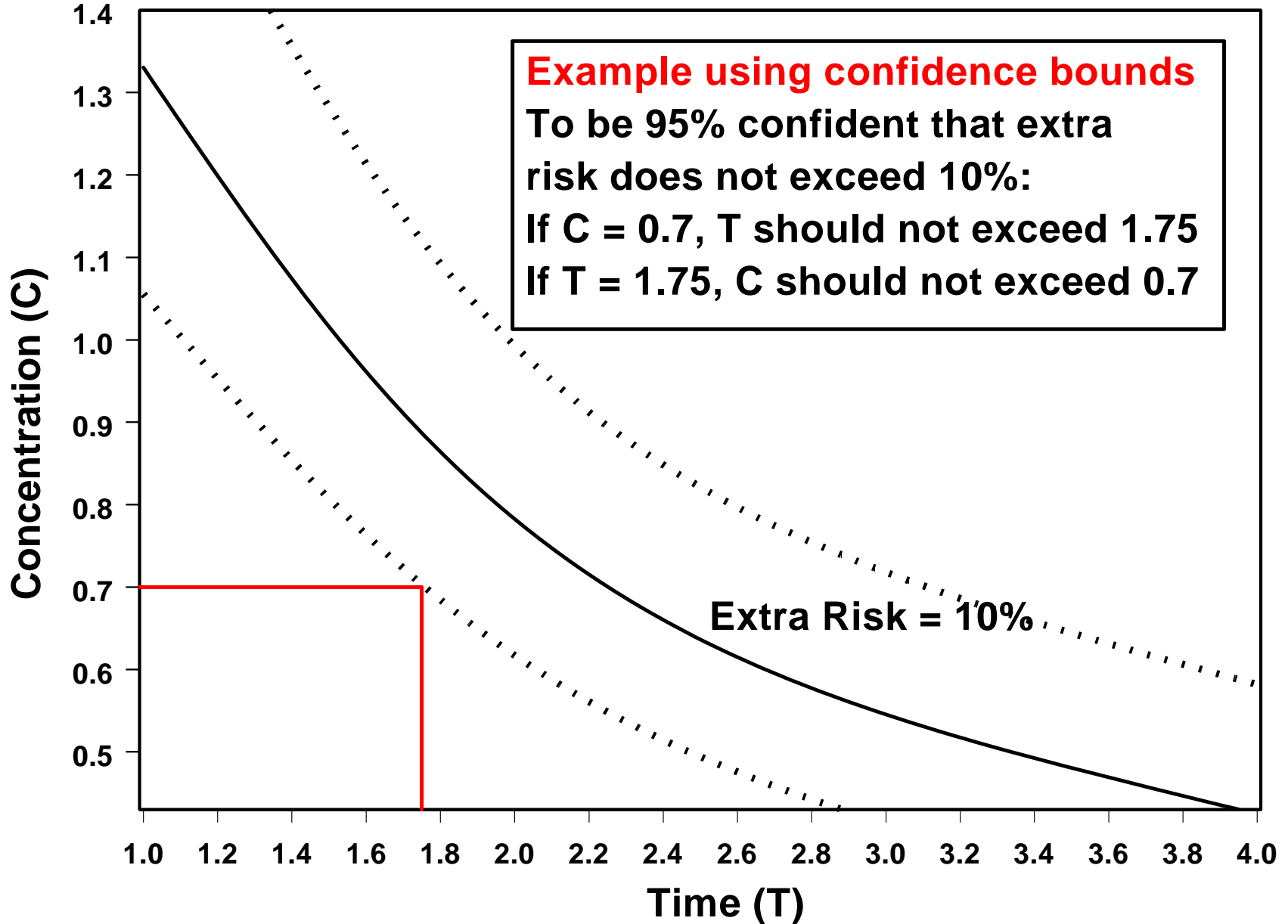
* ERC10 at SEV1

	ERC10	Lower Bound	Upper Bound
1.00 Hours	1.3464	1.0527	1.7220
1.25 Hours	1.1178	0.8835	1.4144
1.50 Hours	0.9602	0.7598	1.2134
1.75 Hours	0.8444	0.6652	1.0718
2.00 Hours	0.7554	0.5906	0.9662

[ec3table() will be revisited]

[Next slide: New – Not in your version of CatReg]

sevtime.csv SEV 1



[Next - test whether Model 2 (unrestricted cumulative odds) can be reduced to Model 1 (cumulative odds)]

> parallel.test()

Hypothesis Test Results

Chisquare	df	p-value
2.197728	2	0.33325

#####

The P Value of the equality test is > 0.05. This is generally considered not significant, indicating that it would be more appropriate to use the simpler restricted cumulative model.

#####

[Next - make the same run again except choose Model 1. Call it Prob 2a]

Problem 2a

(like Prob. 2 except using Model 1, i.e.,
estimate one set of beta parameters instead of two)

Model?

1: cumulative odds model

2: unrestricted cumulative model

Selection: 1 <CR>

AIC : 203.3439 [this compares with AIC = 205.1 for Model 2 above]

Coefficients:

	Estimate	Stan.Err.	Z-Test	p-value	
SEV1	-1.870	0.306	-6.108	1e-05	$[\alpha_1, \text{inter. } 1^{\text{st}} \text{ curve}]$
SEV2	-4.426	0.472	-9.378	1e-05	$[\alpha_2, \text{inter. } 2^{\text{nd}} \text{ curve}]$
LG10CONC	5.405	0.654	8.257	1e-05	$[\beta_1, \text{log(C) coef.}]$
LG10TIME	3.903	0.559	6.979	1e-05	$[\beta_2, \text{log(T) coef.}]$

Generalized R-squared: 0.412 [this compares with 0.419 for Model 2 above]

AIC : 205.0679

Coefficients:

[C-T metric is $\alpha + \beta_1 \log C + \beta_2 \log T$]

	Estimate	S.E.	Z-Test	p-value	
SEV1	-1.957	0.429	-4.557	1e-05	$[\alpha_1, \text{intercept, Sev1}]$
SEV2	-4.695	0.740	-6.344	1e-05	$[\alpha_2, \text{intercept, Sev2}]$
LG10CONC: SEV1	5.230	0.889	5.879	1e-05	$[\beta_{11}, \log(C) \text{ coef., Sev1}]$
LG10TIME: SEV1	4.360	0.864	5.040	1e-05	$[\beta_{21}, \log(T) \text{ coef. , Sev1}]$
LG10CONC: SEV2	6.306	1.206	5.229	1e-05	$[\beta_{12}, \log(C) \text{ coef., Sev2}]$
LG10TIME: SEV2	3.501	0.713	4.908	1e-05	$[\beta_{22}, \log(T) \text{ coef., Sev2}]$

Generalized R-squared: 0.419

[Next – test hypothesis $\beta_1 = \beta_2$. This would provide a further simplification.

> `partest()` <CR> [This is the general command for testing whatever is of interest. Note the similarity of the output format to `parallel.test()`, which is a special case of `partest()`.]

Parameters: SEV1 SEV2 LG10CONC LG10TIME

Enter group of parameters to test for equality (none):

1: LG10CONC (β_1) Equivalent statements of H_0

2: LG10TIME (β_2) $\beta_1 = \beta_2$

3: <CR> $n = 1$, for n in C^nT , since $n = \beta_1/\beta_2$

Test statistics C-T metric [$\alpha + \beta_1\log C + \beta_2\log T$] can be reduced to

Chisquare df p-value [$\alpha + \beta\log(CT)$]

5.363177 1 0.02057

#####

The P Value of the equality test is ≤ 0.05 . This is generally considered significant, indicating that all test parameters should be retained in the model.

#####

[End of monitor display for Prob 2a]

Problem 2a

(Conclusions for Prob 2/2a, for this link function and C-T metric)

- Model 1 is adequate (parallel severity response curves)
- The C-T (or $C \times T$) relationship is estimated as $C^{1.39}T$ (C^nT with $n = \beta_1 / \beta_2 = 5.41 / 3.90 = 1.39$. The form C^nT follows from the choice of C-T metric, i.e., used $\log(c)$ and $\log(T)$)
- Reject the hypothesis that $n = 1$ (i.e, that C-T relationship is CT)
- Do not know at this point if another combination link function and C-T metric might be preferable

Problem 2a

(ERC values for output file, sevtimemod1.out)

* ERC10 at SEV1

	ERC10	Lower Bound	Upper Bound
1 Hours	1.2851	1.0718	1.5409
4 Hours	0.4722	0.3669	0.6079
8 Hours	0.2863	0.2025	0.4047
24 Hours	0.1295	0.0774	0.2165

* ERC10 at SEV2

	ERC10	Lower Bound	Upper Bound
1 Hours	3.8185	3.2065	4.5474
4 Hours	1.4032	1.1945	1.6484
8 Hours	0.8506	0.6601	1.0961
24 Hours	0.3847	0.2515	0.5885

Note:

1. These One-sided 95% lower bound and one-sided 95% upper bound confidence intervals are equivalent to the lower bound and upper bound of two-sided 90% confidence intervals.

Problem 2a (cont'd)

[Next - calculate ERC values for specified times, e.g., 0.25,0.5,0.75,1.0,4.0 hours. Return to CatReg (which has just been minimized). Use command `ecdata()` first if the current settings for severity level, ERC percentile, or percentile for confidence bounds needs to be changed]

```
>ecdata(<CR>]
```

```
Percentile for ERC? (default=10): <CR>
```

```
Severity level for ERC? (default = 1): 2 <CR>
```

```
Percentile for 1-sided upper and lower confidence intervals? (default=95): <CR>
```

```
> ec3table(duration = c(0.25,0.5,0.75,1.0,4.0))
```

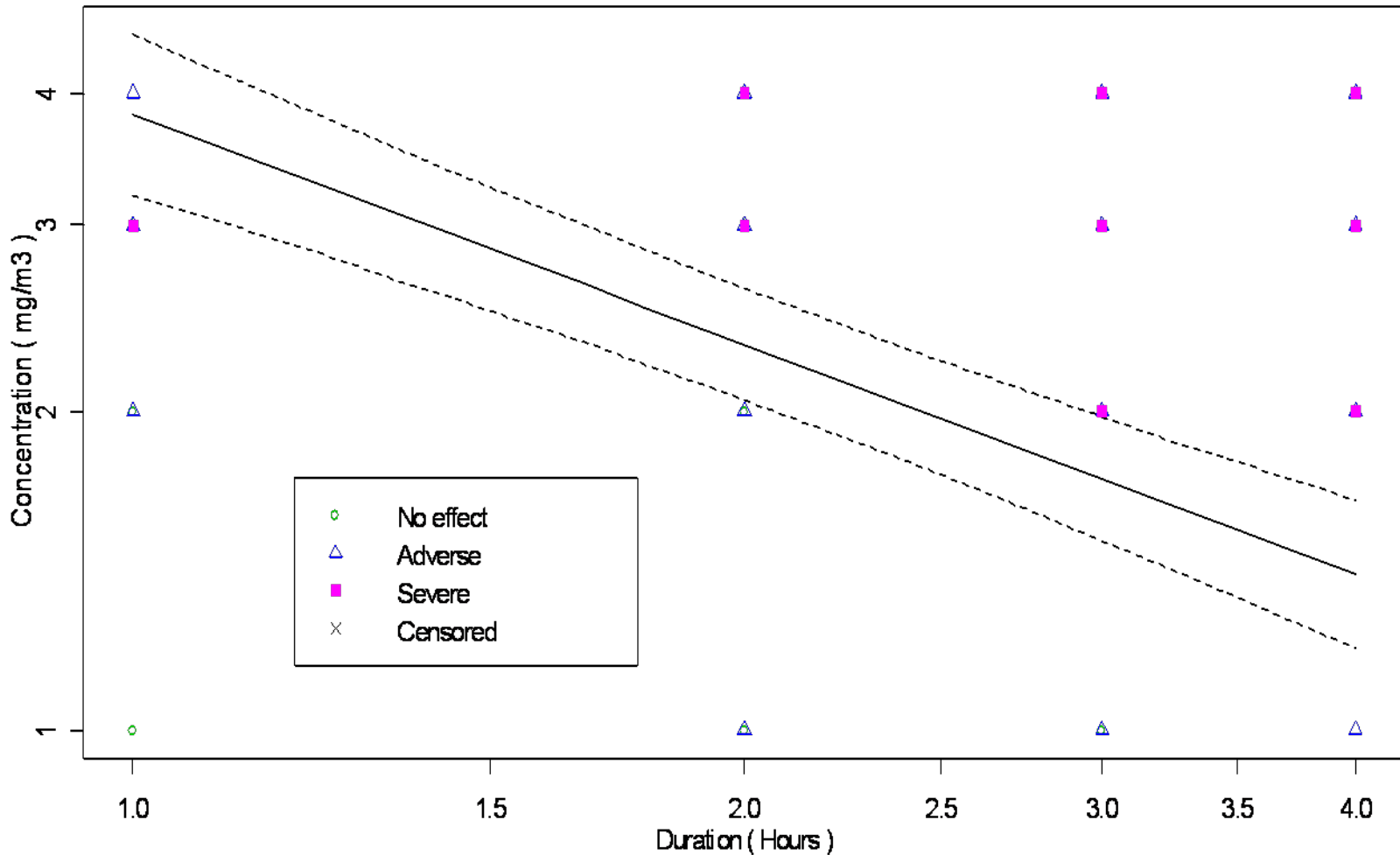
```
* ERC10 at SEV2
```

	ERC10	Lower Bound	Upper Bound of ERC10
0.25 Hours	10.3915	7.1146	15.1778
0.50 Hours	6.2992	4.8037	8.2604
0.75 Hours	4.7003	3.8014	5.8117
1.00 Hours	3.8185	3.2065	4.5474
4.00 Hours	1.4032	1.1945	1.6484

[Next - plot examples. Use `ecdata()` to change current ERC settings as needed]

>catplot(log="xy")<CR> [both axes on log10 scale]

ERC10 Line (SEV2) with 90% two-sided confidence bounds, Link = probit

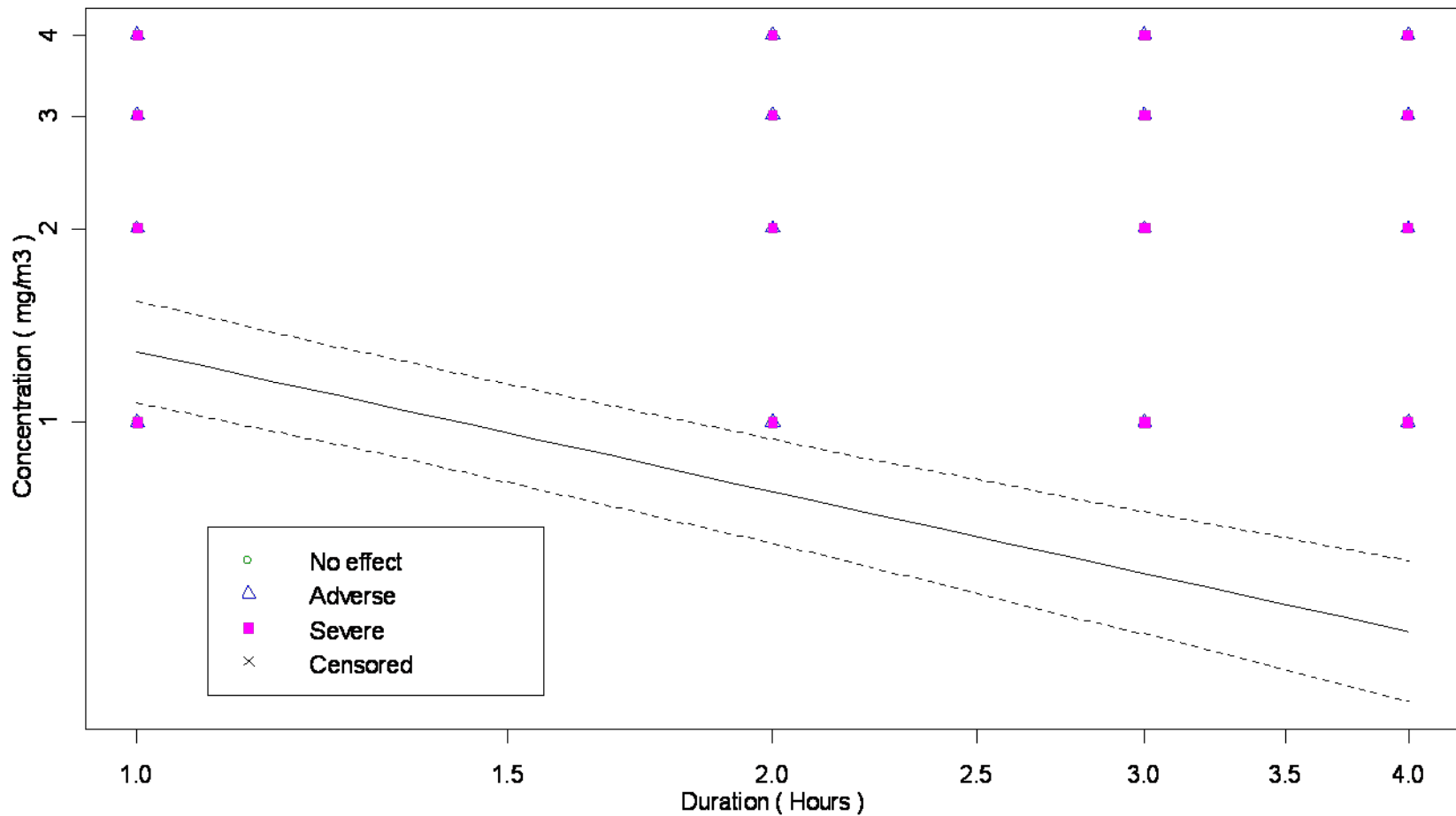


These 90% two-sided confidence bounds are equivalent to 95% one-sided confidence bounds in each direction.

[first used ecdata() to reset severity level to 1]

>catplot(log="xy")<CR> [both axes on log10 scale]

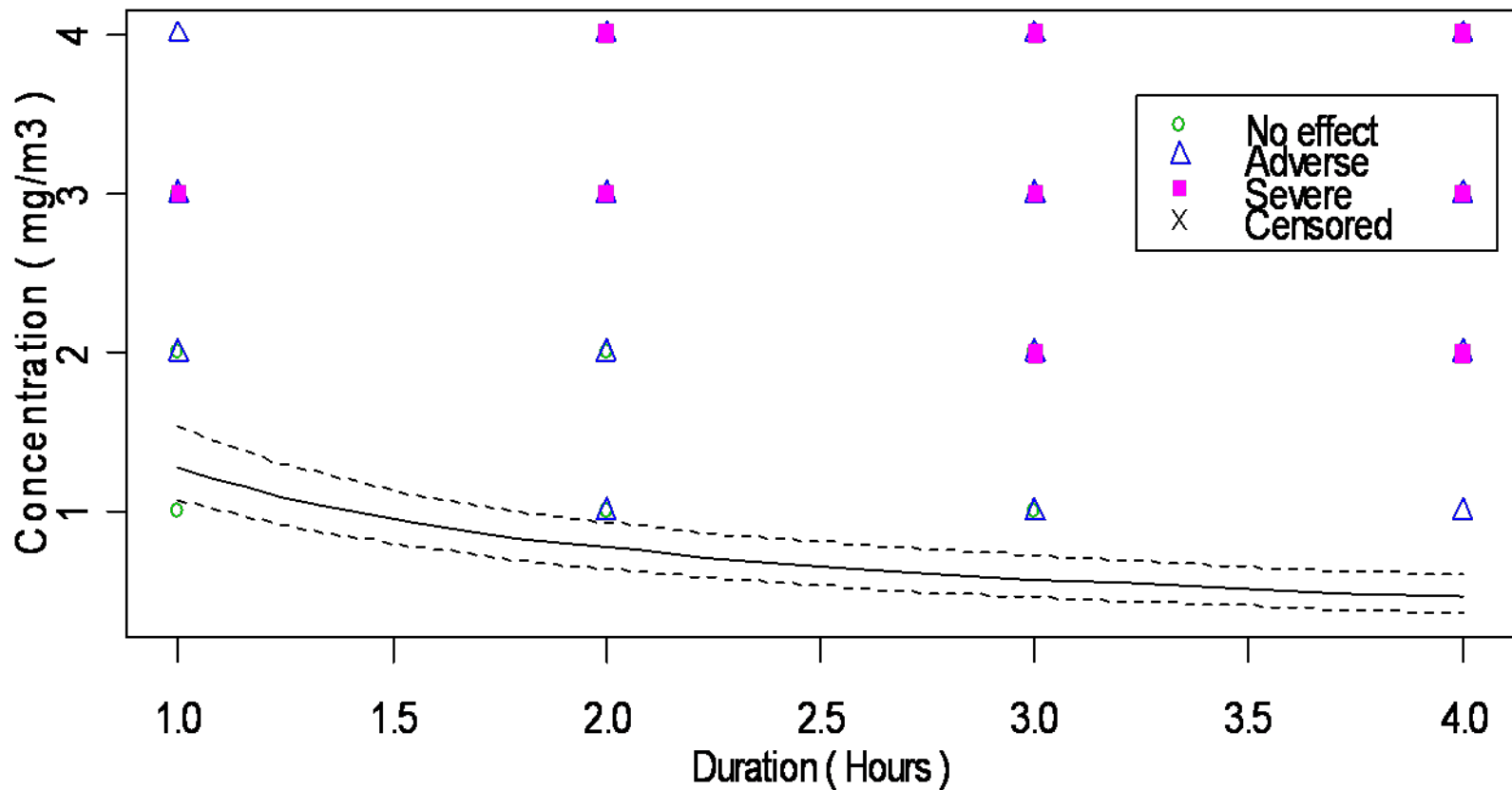
ERC10 Line (SEV1) with 90% two-sided confidence bounds, Link = probit



These 90% two-sided confidence bounds are equivalent to 95% one-sided confidence bounds in each direction.

>catplot(log="")<CR> [both axes on linear scale]

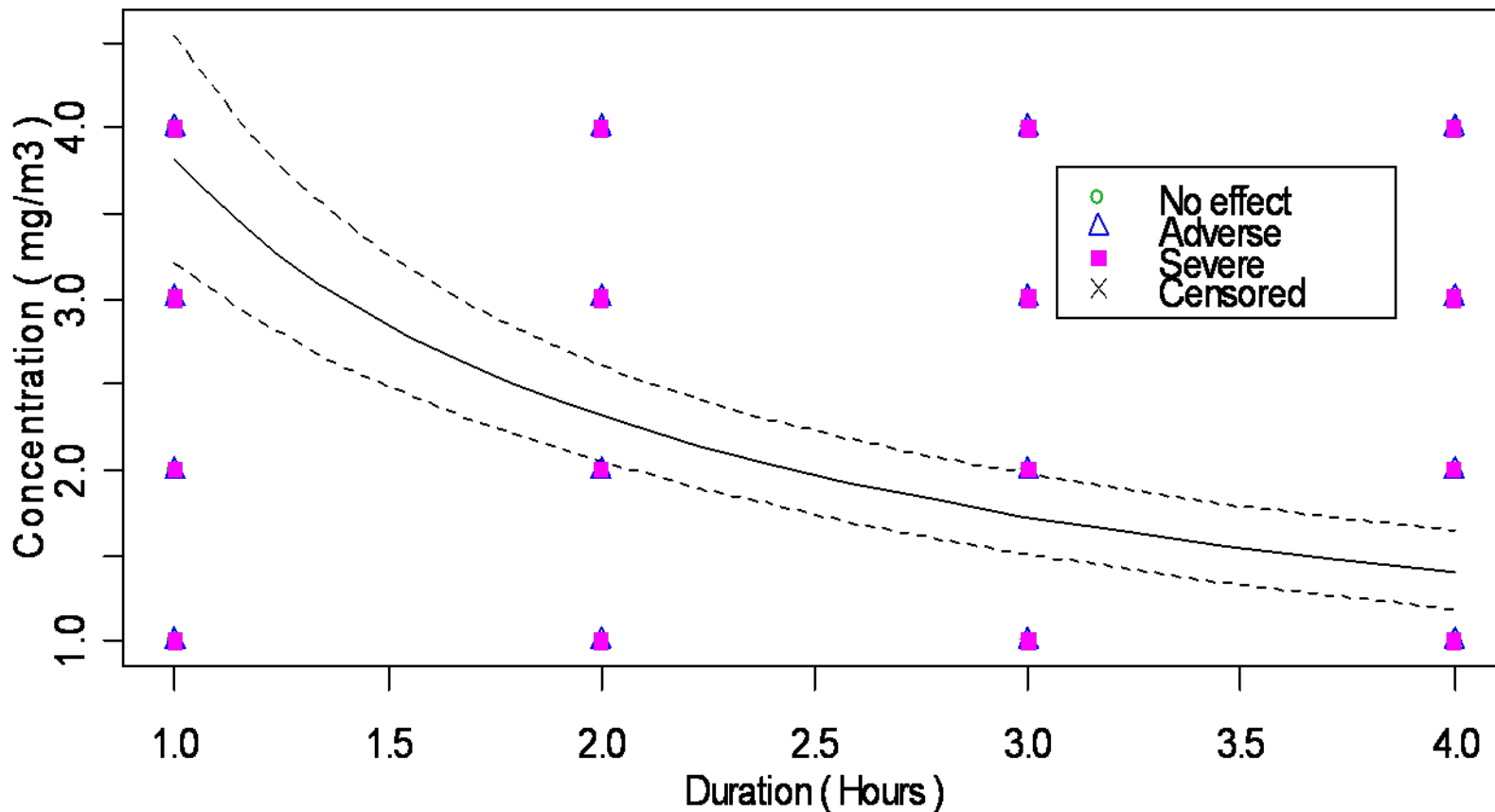
ERC10 Line (SEV1) with 90% two-sided confidence bounds, Link = probit



These 90% two-sided confidence bounds are equivalent to 95% one-sided confidence bounds in each direction.

[first used ecdata() to reset severity level to 2]
>catplot(log="")<CR> [both axes on linear scale]

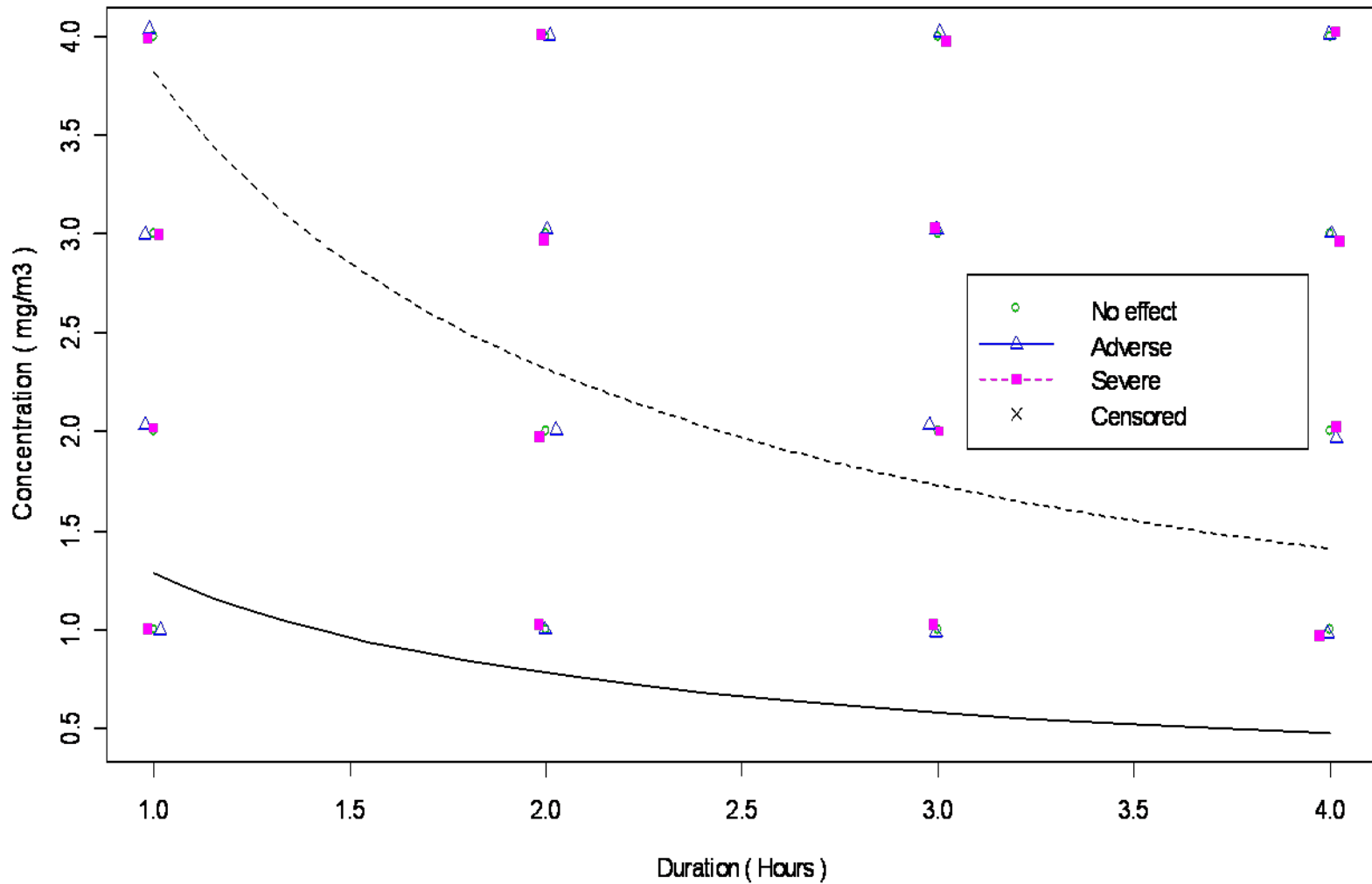
ERC10 Line (SEV2) with 90% two-sided confidence bounds, Link = probit



These 90% two-sided confidence bounds are equivalent to 95% one-sided confidence bounds in each direction.

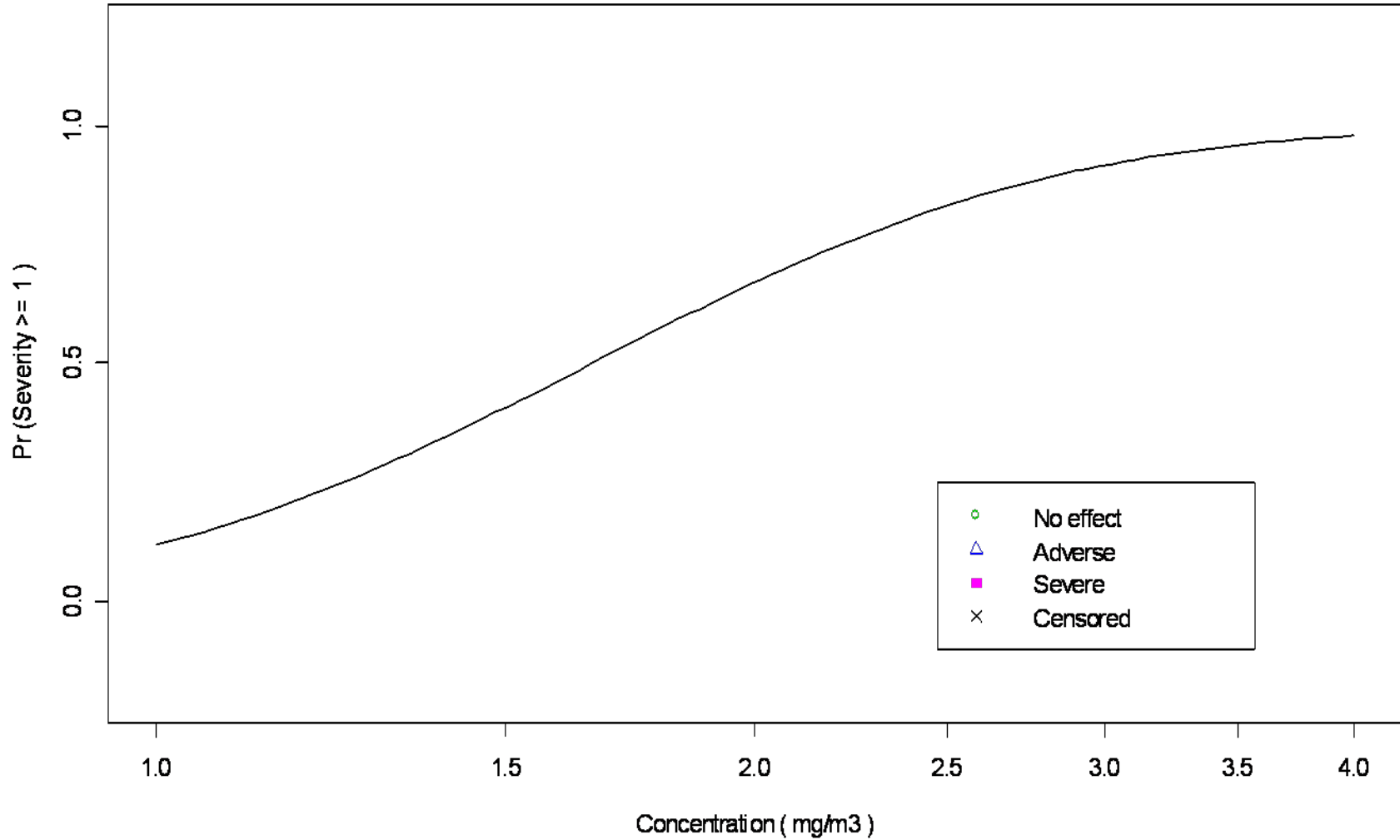
>allsevsplot(log="",jitter = T)<CR> [both axes on linear scale, jittering offsets symbols at the same location] Adverse = Sev1 Severe = Sev2

ERC10 Lines (No Strata) in all SEV levels



```
>prplot(time=1.5)<CR>
```

Duration (Hours) = 1.5 Stratum = SEV1



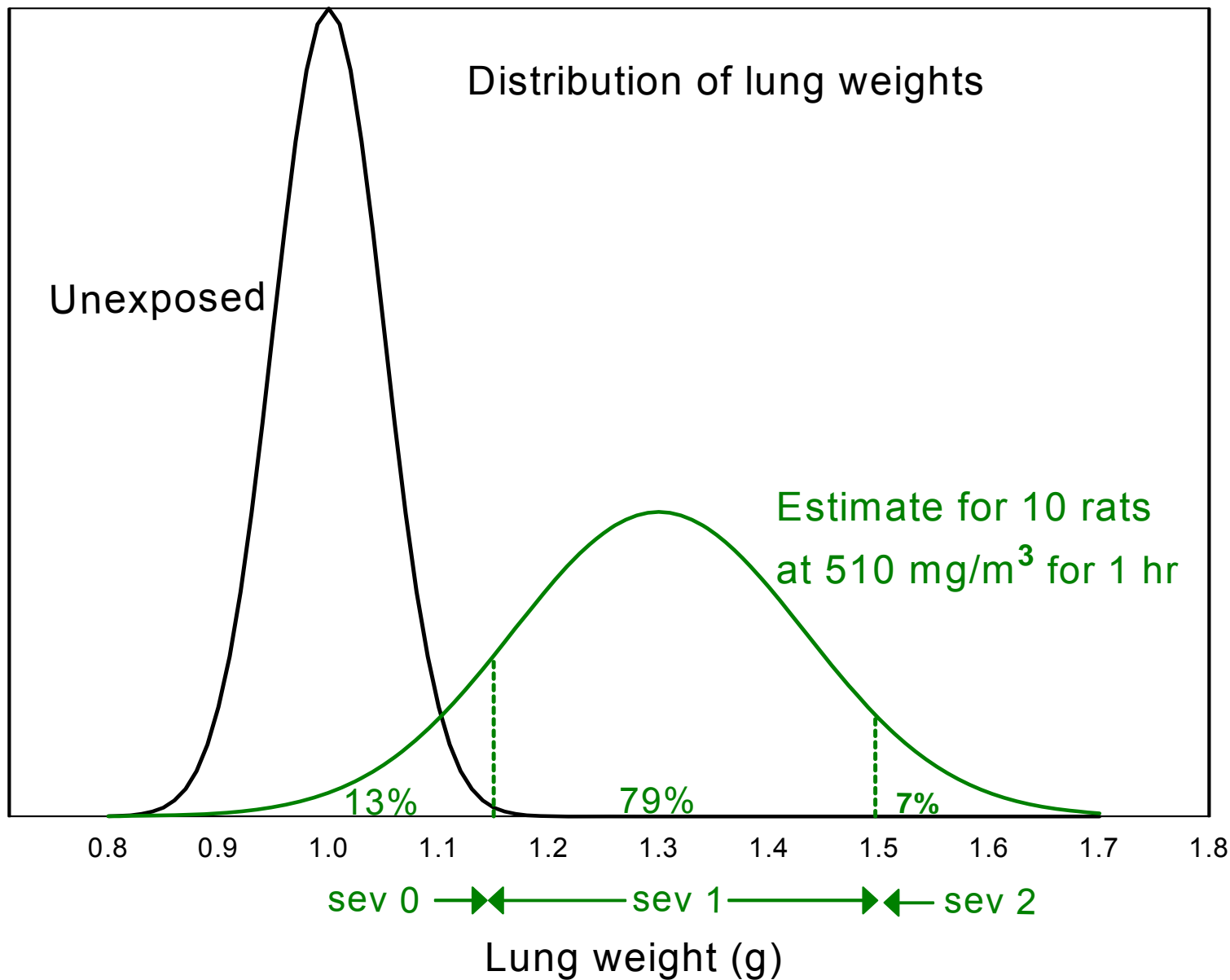
4. Example of categorizing continuous data (variable time, $S \geq 1$)

- Studies report mean and standard error (SE) for each treatment group, e.g., lung weight
- Decide maximum value that would occur in unexposed e.g., if mean = 1.0g, standard deviation = 0.05, then there would be no occurrences above 1.15g
- Set severity levels from intervals of lung weight, e.g.,
 - Sev0 (<1.15g)
 - Sev1 (1.15g-1.50g)
 - Sev2 (>1.50g)

Example of categorizing continuous data (cont'd)

- Assume a distribution for lung weights, e.g., normal
- Estimate proportion of each treatment group in each interval and multiply by sample size (N_{sub}) to get an incidence for each severity level
- Results are usually not whole numbers but total Incid must still equal N_{sub} for each treatment group
- Technical details are in Appendix A of user manual

[Note: If individual measurements of lung weight are available, use them to get an incidence for each severity level. This estimation procedure is only for when just the mean and standard error are reported]



Example of continuous data from User Manual

TABLE A-1. SUMMARY DATA FOR “REF.ID” = 2 IN TABLE 3-6

Index	Exp.	Group	mg/m3	Hours	Avg. Lung Weight (\bar{X})	Std. Error
1	1	1	330	2	1.1	0.03
2	1	2	360	2	1.2	0.025
3	1	3	390	2	1.5	0.04
4	1	4	410	2	1.8	0.08
5	1	5	460	2	1.7	0.1
6	2	1	460	1	1.2	0.025
7	2	2	510	1	1.3	0.04
8	2	3	560	1	1.6	0.08
9	2	4	610	1	1.8	0.1
10	3	1	560	0.5	1.3	0.025
11	3	2	610	0.5	1.5	0.04
12	3	3	660	0.5	1.6	0.08
13	3	4	710	0.5	1.8	0.1

Example of continuous data from User Manual (cont'd)

Input file for summary data in Table A-1

Marker	Ref.id	Exp	Group	Species	Sex	mg/m3	Hours	Endpoint	Nsub	Incid	BestNum	SevLo	SevHi
22	2	1	5	RT	M	460	2	Lung wt	10	1.5	1	1	1
23	2	1	5	RT	M	460	2	Lung wt	10	7.2	2	2	2
24	2	2	1	RT	M	460	1	Lung wt	10	2.9	1	0	0
25	2	2	1	RT	M	460	1	Lung wt	10	7.1	1	1	1
26	2	2	2	RT	M	510	1	Lung wt	10	1.3	0	0	0
27	2	2	2	RT	M	510	1	Lung wt	10	7.9	1	1	1
28	2	2	2	RT	M	510	1	Lung wt	10	0.7	2	2	2
29	2	2	3	RT	M	560	1	Lung wt	10	3.5	1	1	1
30	2	2	3	RT	M	560	1	Lung wt	10	6.5	2	2	2
31	2	2	4	RT	M	610	1	Lung wt	10	1.6	1	1	1
32	2	2	4	RT	M	610	1	Lung wt	10	8.1	2	2	2
33	2	3	1	RT	M	560	0.5	Lung wt	10	9.4	1	1	1
34	2	3	2	RT	M	610	0.5	Lung wt	10	4.9	1	1	1
35	2	3	2	RT	M	610	0.5	Lung wt	10	5.0	2	2	2
36	2	3	3	RT	M	660	0.5	Lung wt	10	3.6	1	1	1
37	2	3	3	RT	M	660	0.5	Lung wt	10	6.5	2	2	2
38	2	3	4	RT	M	710	0.5	Lung wt	10	1.6	1	1	1
39	2	3	4	RT	M	710	0.5	Lung wt	10	8.1	2	2	2

5. Add other variables

- Take into account effect of variables within the data besides severity level, C and T
- Two types of effects
 - Fixed. Interested in estimating and/or testing for differences, e.g., species, strain, gender. Handled in CatReg by “stratification” of parameters, e.g., allowing one or more parameters to vary by species.
 - Random. Interested into taking into account variability between samples, e.g., studies. Studies would be treated as a random sample of possible studies. Parameter estimates are not affected, but confidence intervals are wider due to inter-sampling variability. Handled by “clustering” (to be covered in next class).

Problem 3

(description)

- This is a hypothetical problem conceptually similar to Problem 2
- Differences are two species (MU = mouse, RT= rat), response in two organs (L = lung, C = CNS), and two experiments (Exp)
- Each test animal was examined after exposure by inhalation for adverse effects on the lungs and CNS, and classified as having “no effect”, “small effect”, or “large effect” separately in each both organ (S = 2)

Problem 3

(objective)

- Analyze CNS response only (filter out lung response)
- Stratify intercept, concentration, and duration on species (creates separate parameters $(\alpha, \beta_1, \beta_2)$ for mice and rats)
- Test where differences lie between mice and rats (coefficients for intercept, concentration, duration [time])
- Simplify stratification as possible and run CatReg again if needed
- Evaluate difference in ERC10 between rats and mice

Problem 3

(objective – cont'd)

- Use $\log(C)$, $\log(T)$ and logit link
- Use Model 1 (only α differs by severity level, but separately for mice and rats)
- Assume negligible difference between Exp. 1 and Exp. 2, e.g., due to different materials and methods

Problem 3

(first part of input file, chemx.csv)

Exp.	Group	Species	Target	mg/m3	Hours	SevLo	Nsub	Incid
1	1	MU	C	1259	1.25	0	10	10
1	2	MU	C	1259	1.6	0	10	9
1	2	MU	C	1259	1.6	1	10	1
1	3	MU	C	1259	2	0	10	4
1	3	MU	C	1259	2	1	10	6
1	4	MU	C	1259	2.5	0	10	1
1	4	MU	C	1259	2.5	1	10	9
1	5	MU	C	1585	1.25	0	10	7
1	5	MU	C	1585	1.25	1	10	3
1	6	MU	C	1585	1.6	0	10	3
1	6	MU	C	1585	1.6	1	10	7

Exp (1,2) Species (MU,RT) Target (C,L)

Problem 3

(monitor display)

[Filter data on lung endpoint]

Variables:

Exp. Group Species Target mg/m3 Hours SevLo Nsub Incid Incid GpSize

Filter the data on which variables? (none):

1: **Target**

Categories of Target:

C L

Categories to exclude? (none):

1: **L**

[Stratify all parameters on species]

Variables:

Exp. Group Species Target mg/m3 Hours SevLo Nsub Incid Incid GpSize

Stratify intercepts on which variables? (none): [stratify α]

1: **Species**

Stratify concentration (mg/m3) on which variables? (none): [stratify β_1]

1: **Species**

Stratify time (Hours) on which variables? (none): [stratify β_2]

1: **Species**

[summary]

Input file : chemx.csv

Filtered data: Target : L

Model : cumulative odds model

Link : logit

Clustering : none

Message :

Iterations : 54 19

Deviance : 362.2864

Residual DF : 57

AIC : 376.2864 [Note AIC = 376.3]

Scale:

Concentration: log10(mg/m3)

Duration : log10(Hours)

Stratification:

Intercept : Species

Concentration: Species

Duration : Species

[To interpret intercept terms: Sev1 and Sev2 are α_1 and α_2 for MU (mice). To get Sev1 and Sev2 for RT(rats), add RT:INTERCEPT to corresponding term for MU, e.g., Sev1 for RT = Sev1 for MU + RT:INTERCEPT = -63.41 - 18.88. This assumes that the difference in the intercepts between rats and mice are the same for both severity levels.]

Coefficients:

	Estimate	Std. Error	Z-Test=0	p-value
SEV1	-63.41411	7.416329	-8.550606	0.00001
SEV2	-67.73756	7.628362	-8.879699	0.00001
MU:INTERCEPT	0.00000	0.000000	NA	NA
RT:INTERCEPT	-18.87883	9.405362	-2.007241	0.04472
MU:LG10CONC	18.76326	2.224125	8.436242	0.00001
RT:LG10CONC	24.44292	2.569453	9.512886	0.00001
MU:LG10TIME	18.44227	2.213848	8.330411	0.00001
RT:LG10TIME	16.47041	2.198142	7.492878	0.00001

Generalized R-squared: 0.474

#####

Note: About 47.4 % of the variation in the response is accounted for by the explanatory variables in the current model fit.

The p-value of the model fit is ≤ 0.05 . This is generally considered significant, indicating that the current model fit is acceptable.

#####

[Test hypothesis that mouse and rat curves are not significantly different]

> partest()

Parameters:

SEV1 SEV2 RT:INTERCEPT MU:LG10CONC RT:LG10CONC MU:LG10TIME
RT:LG10TIME

Enter parameters to test for removal (none):

1: RT:INTERCEPT

Enter group of parameters to test for equality (none):

1: MU:LG10CONC

2: RT:LG10CONC

Enter another group to test for equality (none):

1: MU:LG10TIME

2: RT:LG10TIME

Hypothesis Test Results

Coefficients tested for removal: RT:INTERCEPT

Coefficients tested for equality:

Group 1 : MU:LG10CONC RT:LG10CONC

Group 2 : MU:LG10TIME RT:LG10TIME

Test statistics:

Chisquare	df	p-value	
15.85930	3	0.00121	[Conclude that response curves are different for MU and RT. Next - which parameters are different?]

Prob. 3 Results

Prob 3. chemx.csv RT(rat) and MU(mouse), C(CNS), log(C), log(T), logit link, Model 1						
MU:INTERCEPT=	MU:LG10CONC=	MU:LG10TIME=	TEST			MODEL FIT
RT:INTERCEPT	RT:LG10CONC	RAT:LG10TIME	P-VALUE	AIC	R2	P-VALUE
				376.3	0.474	0
X	X	X	0.001			
X			0.045			
	X		0.043			
		X	0.472	374.8	0.473	0
X	X		0.127	376.5	0.467	0
X		X	0.034			
	X	X	0.043			

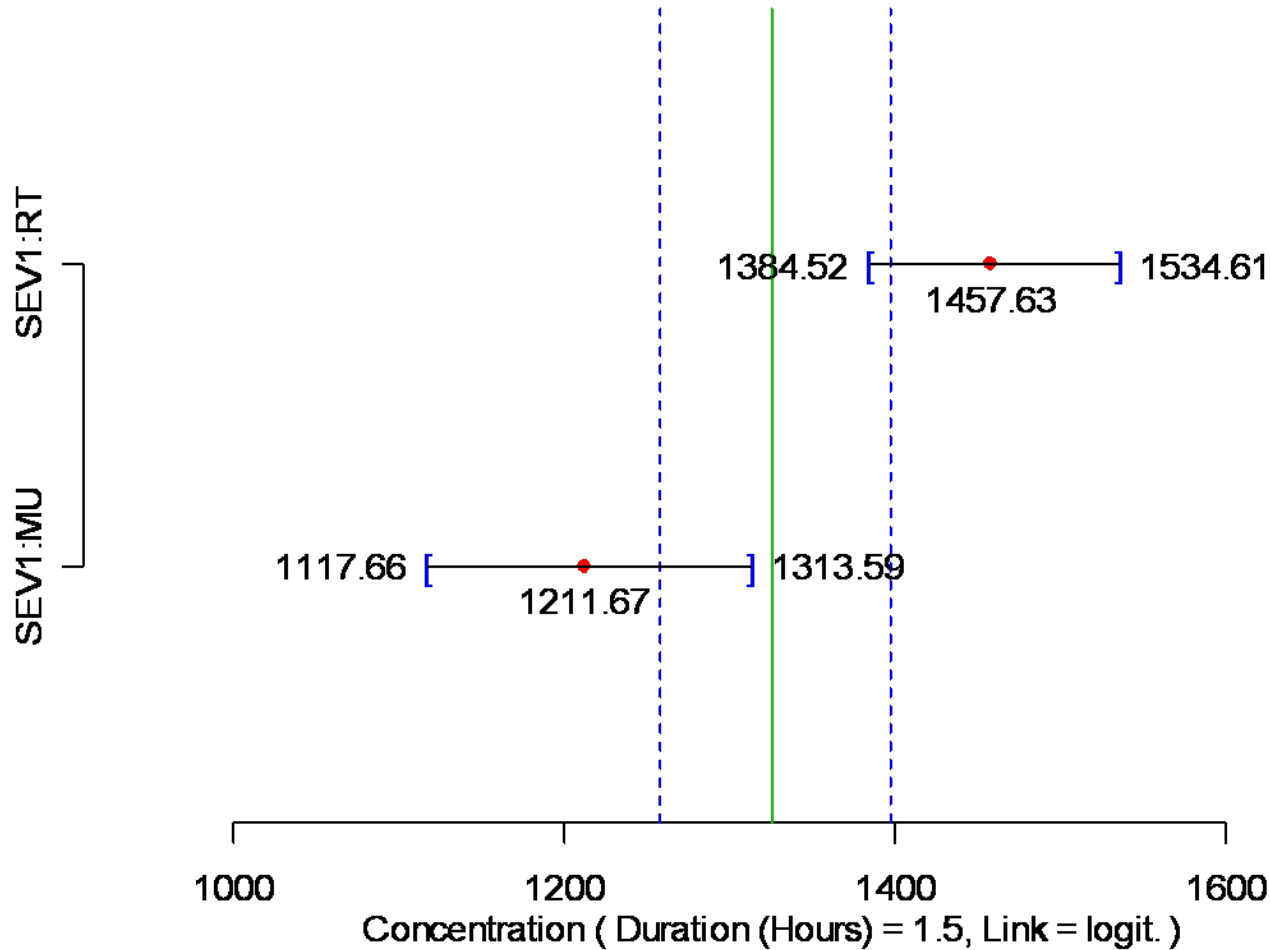
Conclusion - Rats and mice have different intercepts and coefficients of concentration but no evidence of different coefficients of time

[Next - plots of ERC10 for Sev1 for highlighted case]

>confplot(time=1.5) [Sev1 appears because that is the last setting]

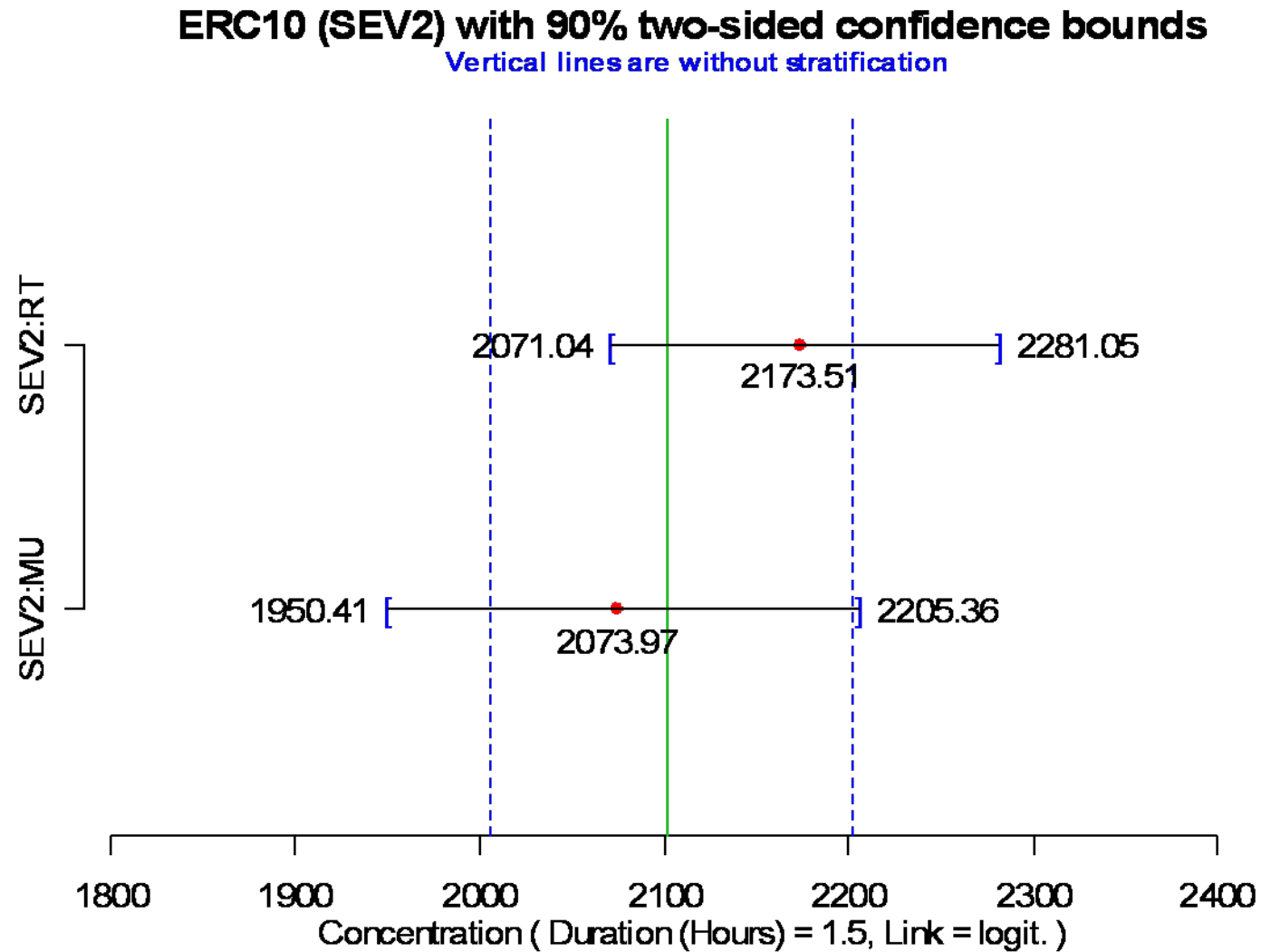
ERC10 (SEV1) with 90% two-sided confidence bounds

Vertical lines are without stratification



These 90% two-sided confidence bounds are equivalent to 95% one-sided confidence bounds in each direction.

>confplot(time=1.5) [ecdata() was used first to change severity level to Sev2]



These 90% two-sided confidence bounds are equivalent to 95% one-sided confidence bounds in each direction.

Some questions we could address

(using run for “best” outcome in the results table)

- Do rats or mice appear to be more sensitive at severity 2?
(Find and compare ERC10 for rats and mice)
- Estimate “n” in C^nT separately for rats and mice . Does severity level matter?
(Recall $n = \beta_1/\beta_2$. Use R as a calculator, e.g. $>3/2<CR$).
Severity level does not matter because we assumed Model 1)
- Can you test the hypothesis that n is equal for rats and mice for severity level 2 in this problem?
(Model assumes β_2 is the same for rats and mice, so use partest to test hypothesis that β_1 is the same)

The End