More treatment levels do increase robustness – an Urban Legend?

Heino Christl[1] & Anja Russ[1]

SETAC EU 2021[2]

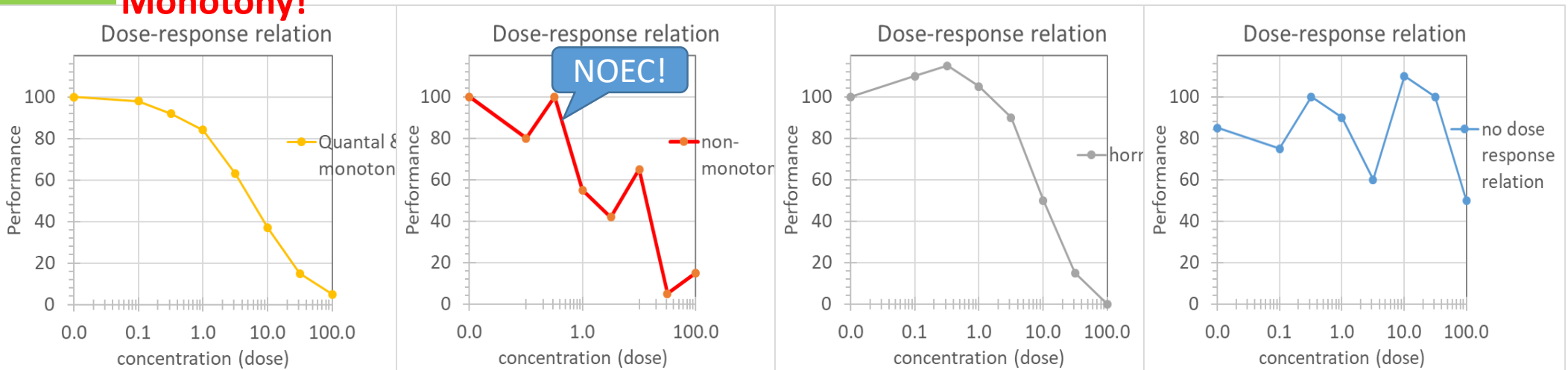
[1]tier3 solutions GmbH, Kolberger Str. 61-63, 51381 Leverkusen, DE

[2]Presentation id: 4636 - Session-id 6.03.06

More dose levels = better curve fit - an urban legend?

- Mesocosm designs ca 2000: → Proposal
  - No replication, but many dose levels (NOEC can't be determined)
  - Expectation: Curve fit will be more robust
- New wave: EFSA's Recurring Issues 2019: EC_{20} or EC_{10} depending on LL+UL
  - Modelled data – varying tested range, no. of dose levels, no. of replicates, at a fixed number of available test systems
  - Hypothesis: More robust curve if more dose levels (and fewer replicates?)
  - Anti-thesis: Narrow spaced dose levels = more often non-monotonous pattern

Monotony!



Three individual questions:

1. **ECx-estimates more robust, if more dose levels?**
   (total no. of test systems unchanged, i.e. fewer/no replicates)
   - Method: Defined dose-response relation with added random scatter
   - Expectation: More dose levels = curve fit will be more robust

2. If EC_{10} or EC_{20} wanted (instead of NOEC+ LOEC), then test **narrow range** (only covering **0** to **20%** effect)?
   - Method: Modelled data – varying tested range (fixed total no. of test systems)
   - Expectation: Narrow test range = more certain EC_{10} or EC_{20}

3. If **only 0%** or **100% effect**, repeat with **additional intermediate test levels**? (until at least two levels with partial effects)?
   - Method: Deterministic survival data – varying individual mortality
   - Expectation: Additional test levels = More precise results (EC_{10} or EC_{50})

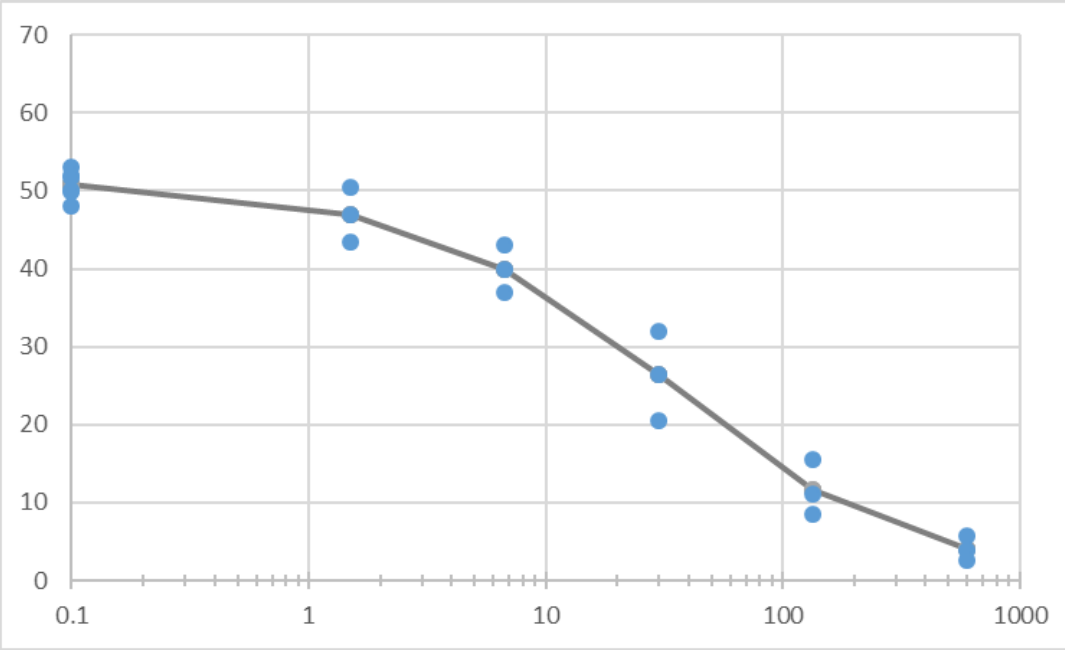
1 Question – more doses, less replicates

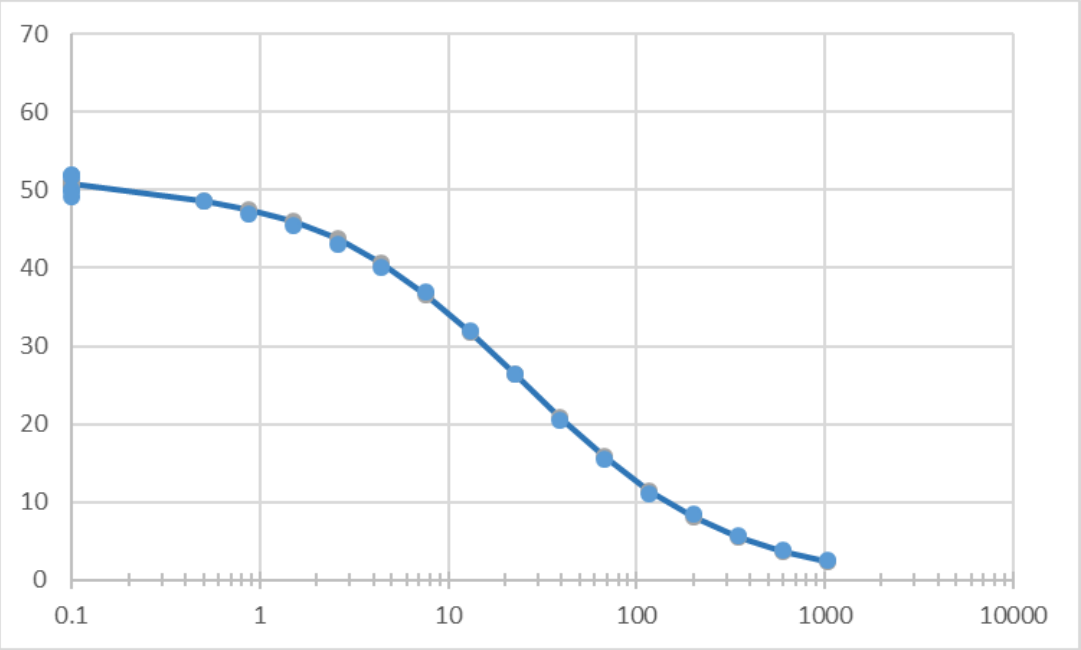
5 dose levels = moderate curve fit

- Scatter of replicates reduces robustness of curve fit

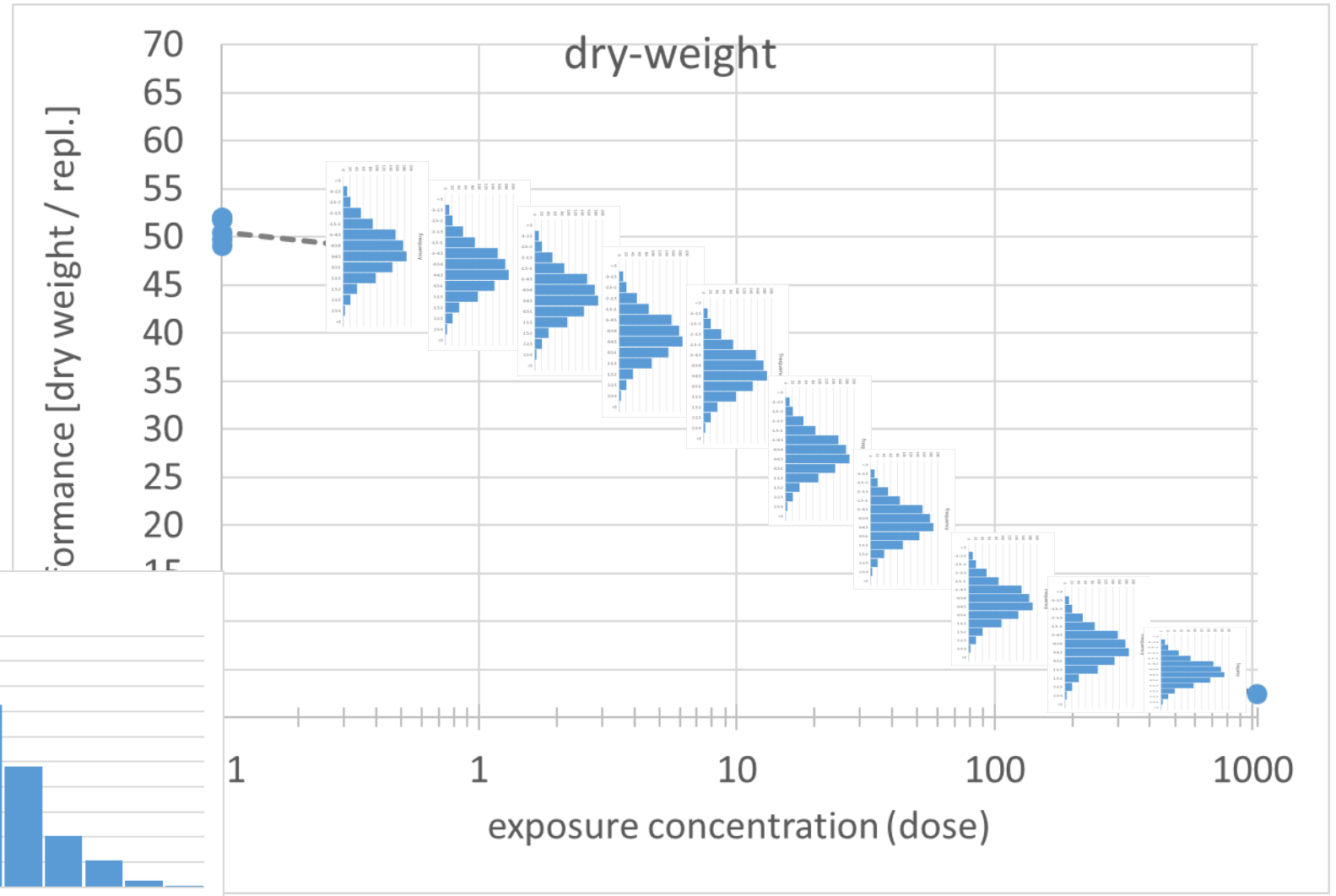
More dose levels = better curve fit?

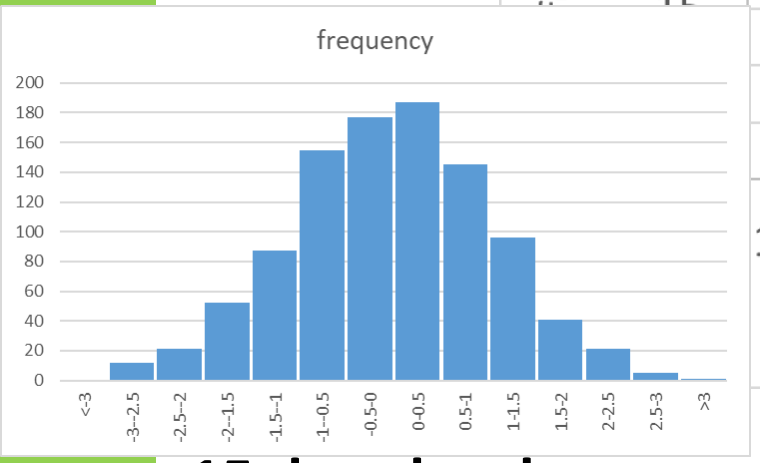
- Hypothesis: No Scatter of replicates = better curve fit + narrow CI





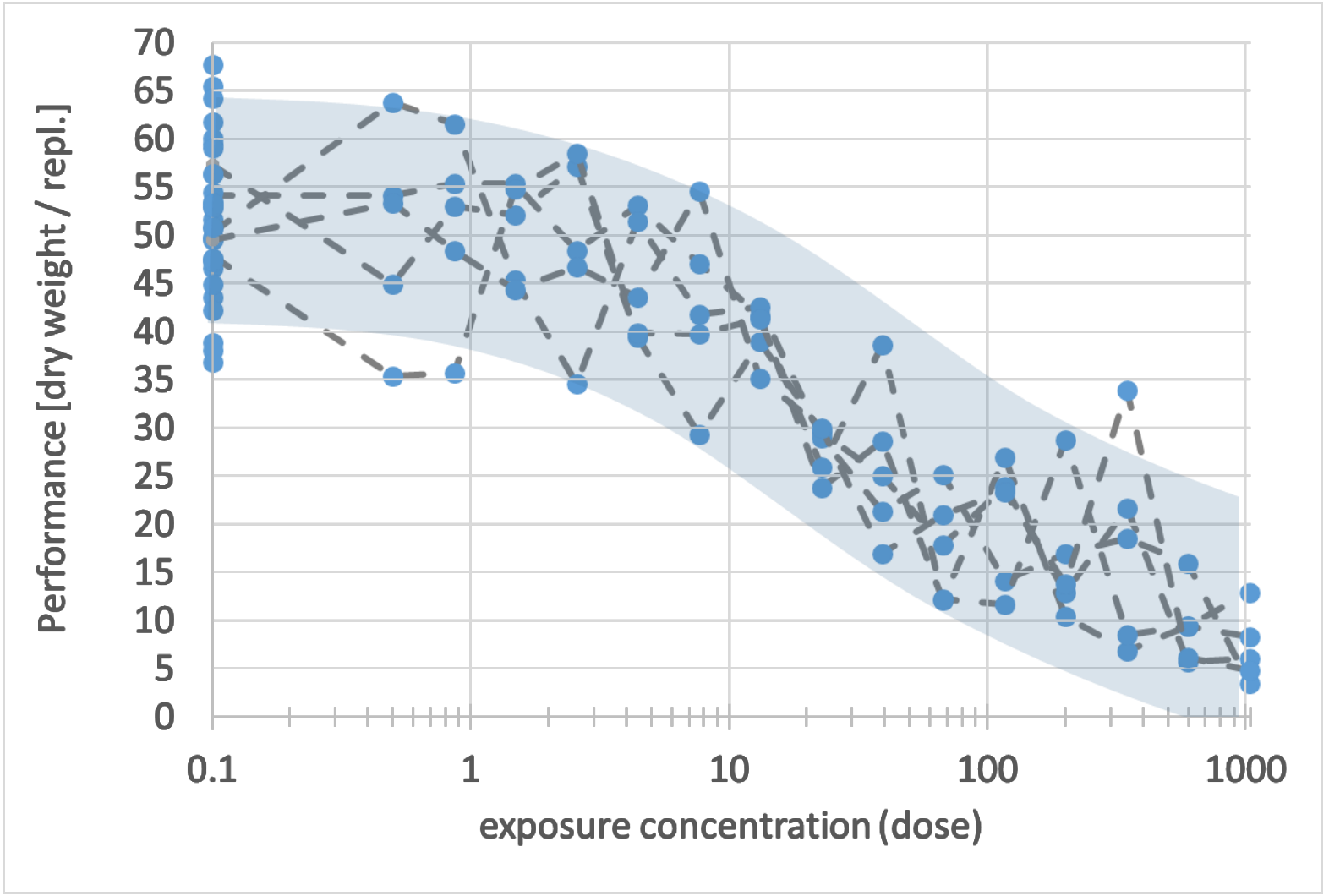
Ideal model - no variability:





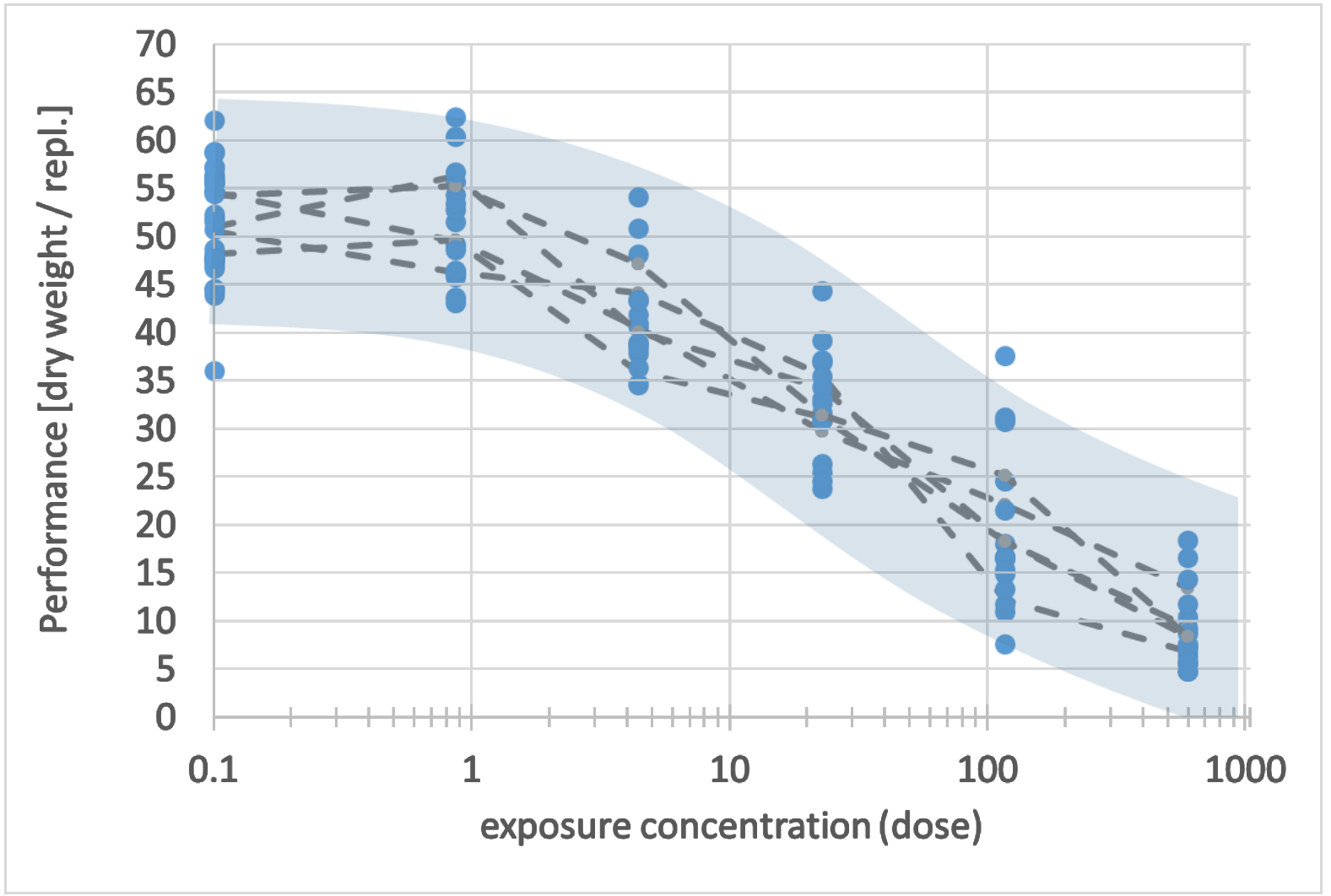
15 dose levels, no replicates (except controls), but normal random scatter

Model random variability:



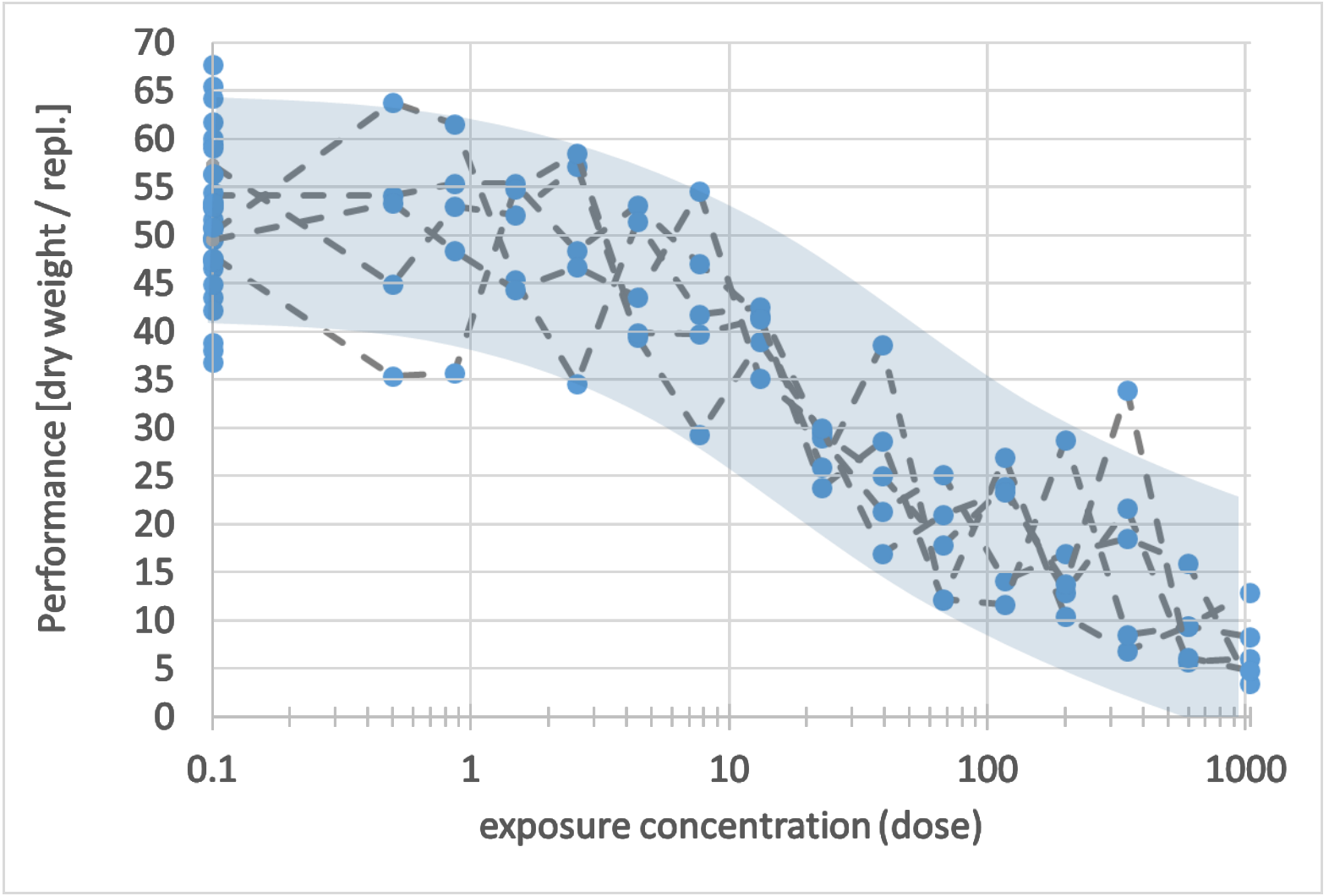
15 dose levels, no replicates (except controls), but normal random scatter

Model random variability:



5 dose levels, three replicates (six controls), same normal random scatter

Model random variability:

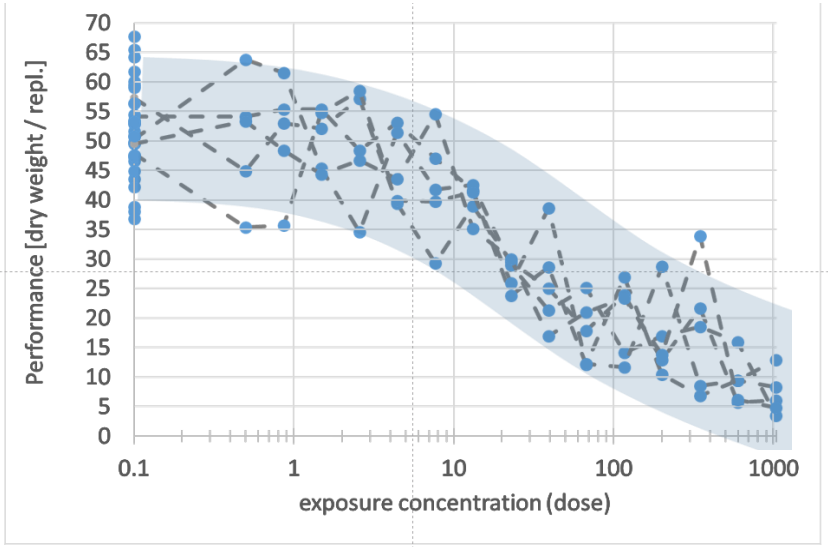


15 dose levels, no replicates (except controls), normal random scatter

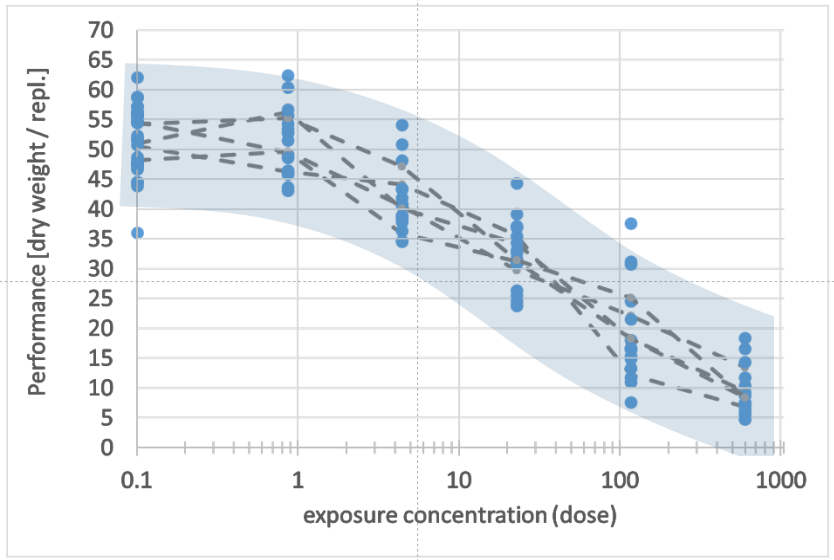
Conclusion 1:

As long as the number of test systems and the variability is the same, it <u>does not matter</u> for the curve fit whether the test systems are used

- for generous replication of few dose levels, or
- for more dose levels with less replication.
- With fewer test systems less certainty, but more replication gets better estimates per dose levels: Less scatter between means (dotted lines right)



15 dose levels, no replicates (except controls), normal random scatter



5 dose levels, three replicates (six controls), same normal random scatter

- What about confidence intervals?

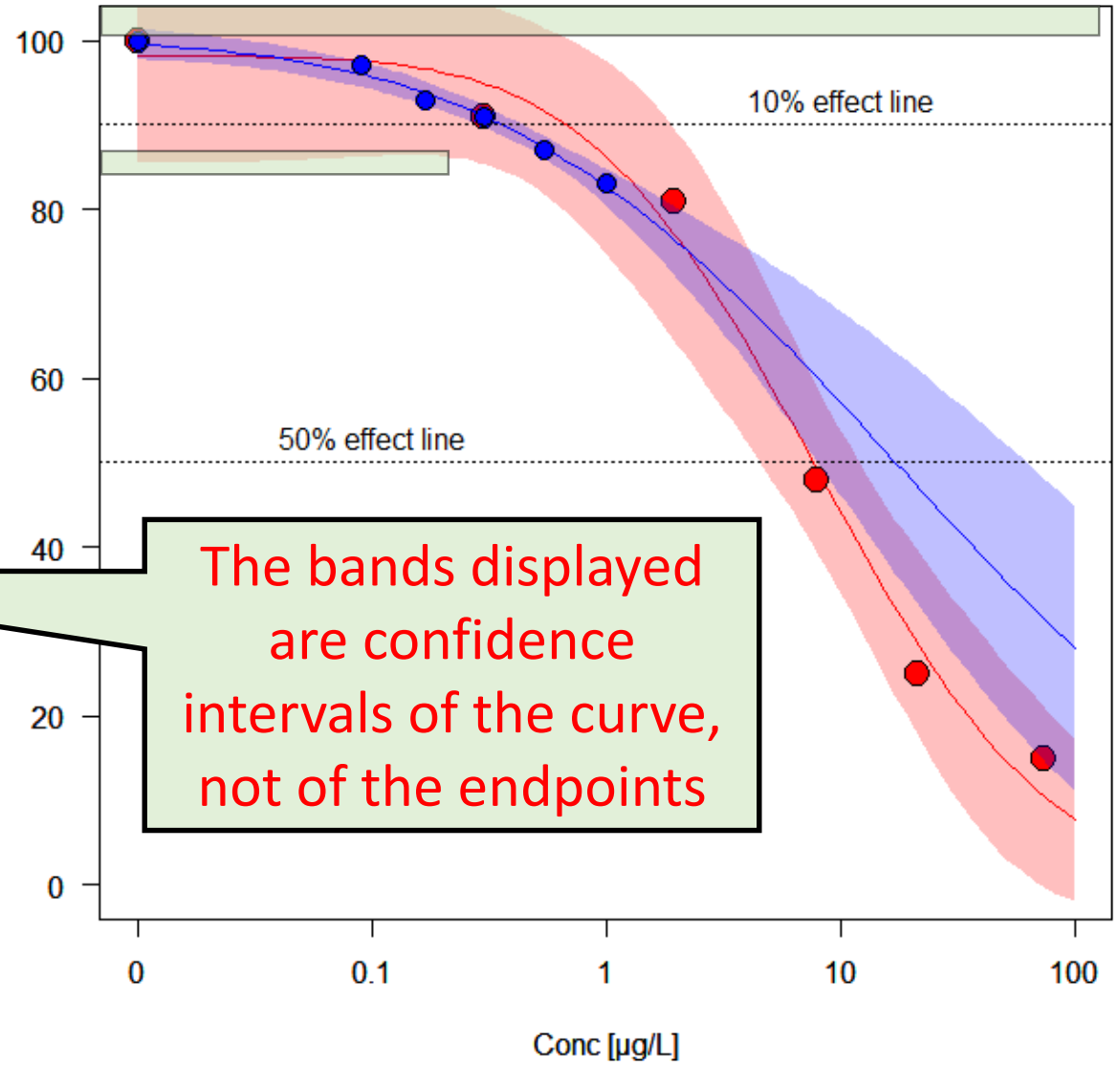
2 Proposed narrow test range – much better confidence-intervals?

Outcome of pesticides peer review meeting on recurring issues in ecotoxicology

EFSA 2009, p 60, (above Fig. F8):
"However, blue points are clustered within a concentration range that yields between 0 and 20% effects. Red points are more spaced, and the highest concentration yields up to 75% effect. Blue points can describe better the dose-response curve between 0% and 20% effect."

"Indeed confidence interval around EC10 is very narrow compared to the confidence provided by the red points. However, at higher effect levels, the confidence provided by the blue points is rapidly decreasing and any ECX with X>20 is just a ventured guess which is likely not representing a reliable value (e.g. see the EC50 illustrated in Figure F8)"

EFSA 2019 "Recurring issues"… ecotox: Appendix E



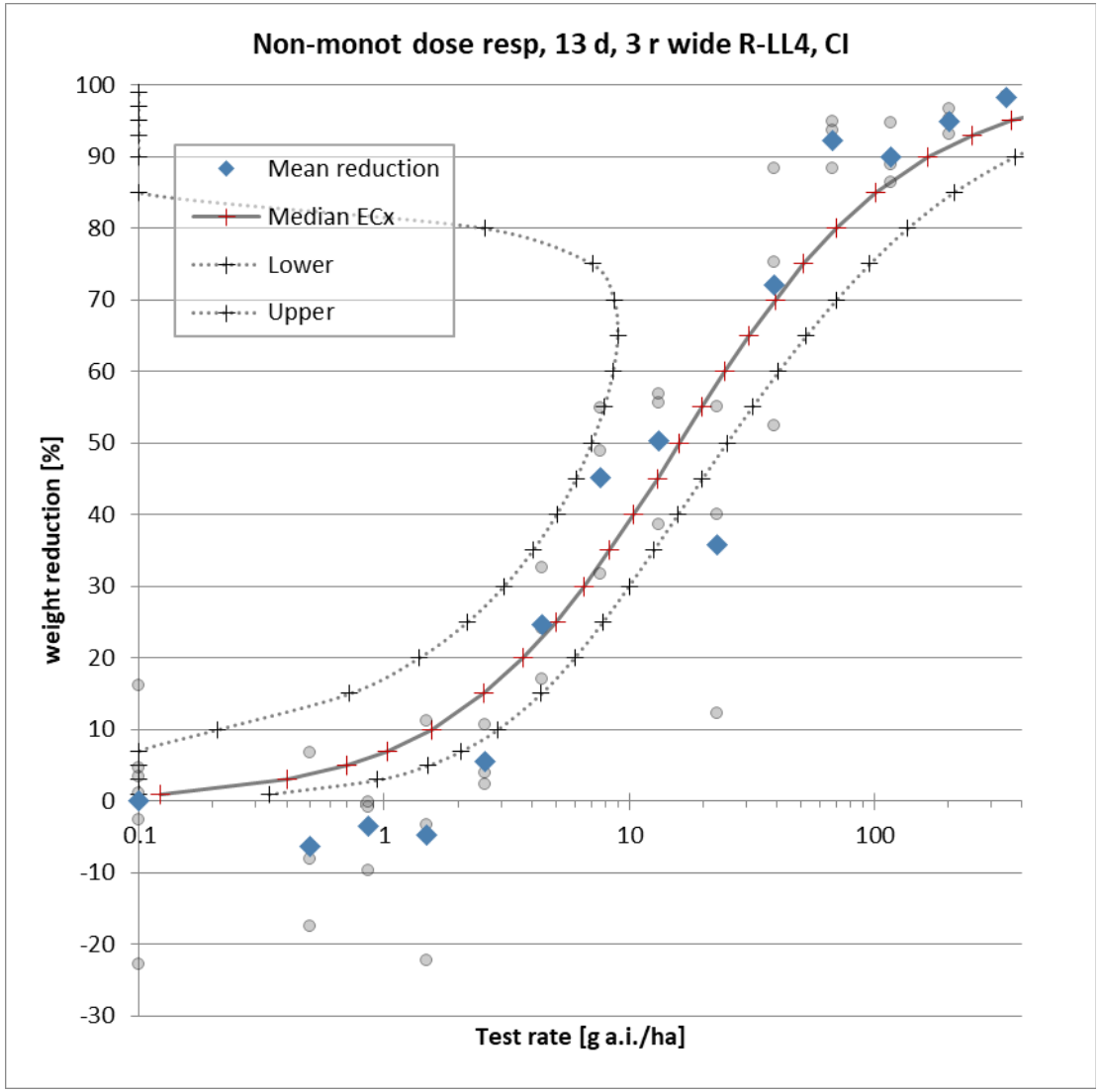
2 Pre-defined data (fixed scatter)

Artificial data:

- **Exact dose-response pattern known (we play God).**
- **Scatter fix (always the same standard deviation)**
- **Distribute artificial data to different number of treatments, or to more or less replicates**
- **Watch how robustness of fit and CI changes**

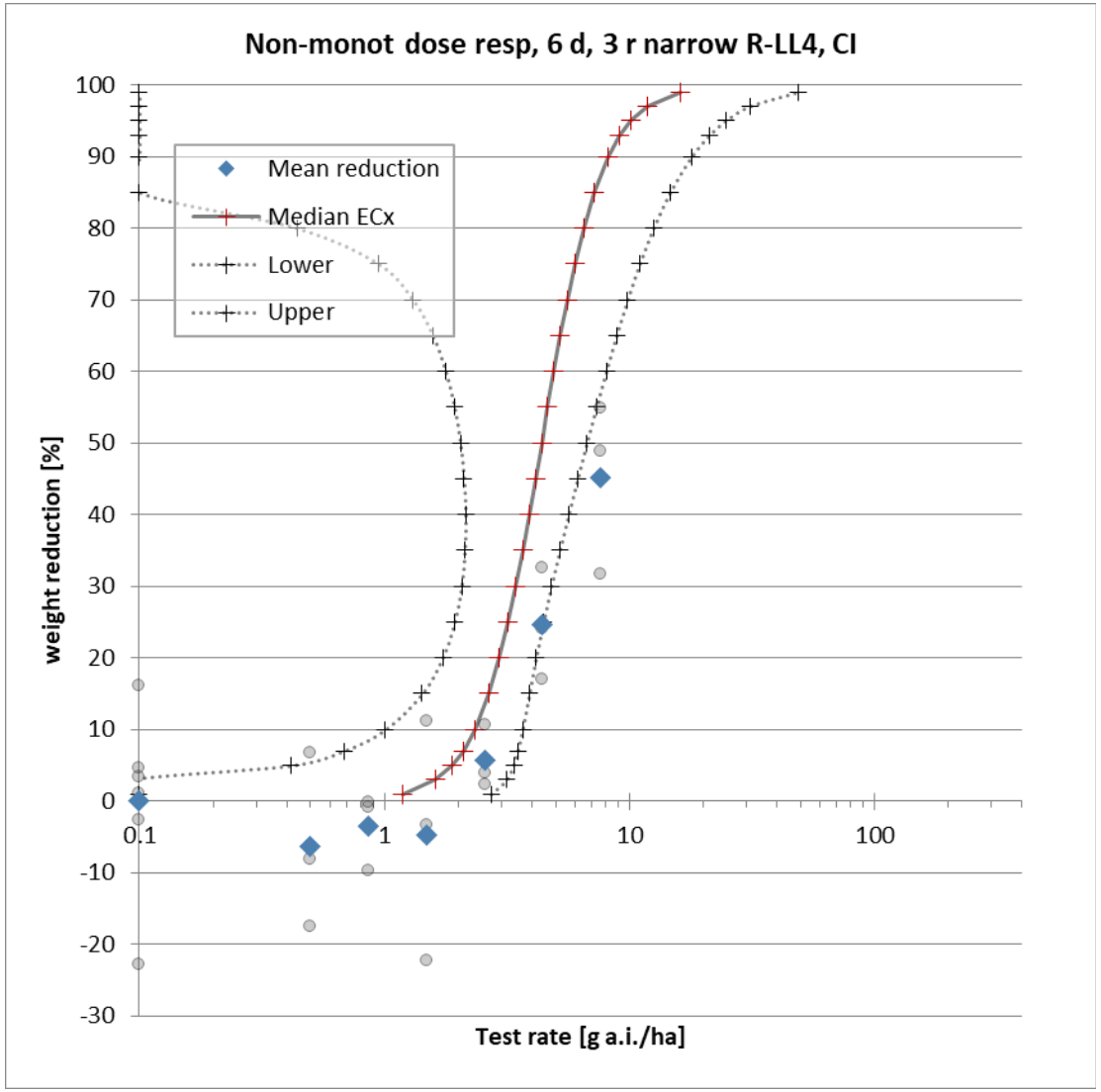
Confidence-intervals of ECx

- ECx around EC10 and EC20 narrow if only that narrow range tested, (6 dose levels, 18 treated systems) but not that much better
- Wide test range with 13 dose levels (39 treated systems) not distinctly better than wide test range with 5 dose levels (15 treated systems)
- With narrow range risk distinctly increased to obtain uninterpretable data “CI not determined for mathematical reasons” (e.g. negative estimates)
- Wide range with 5 dose levels almost as good as with 13 dose levels



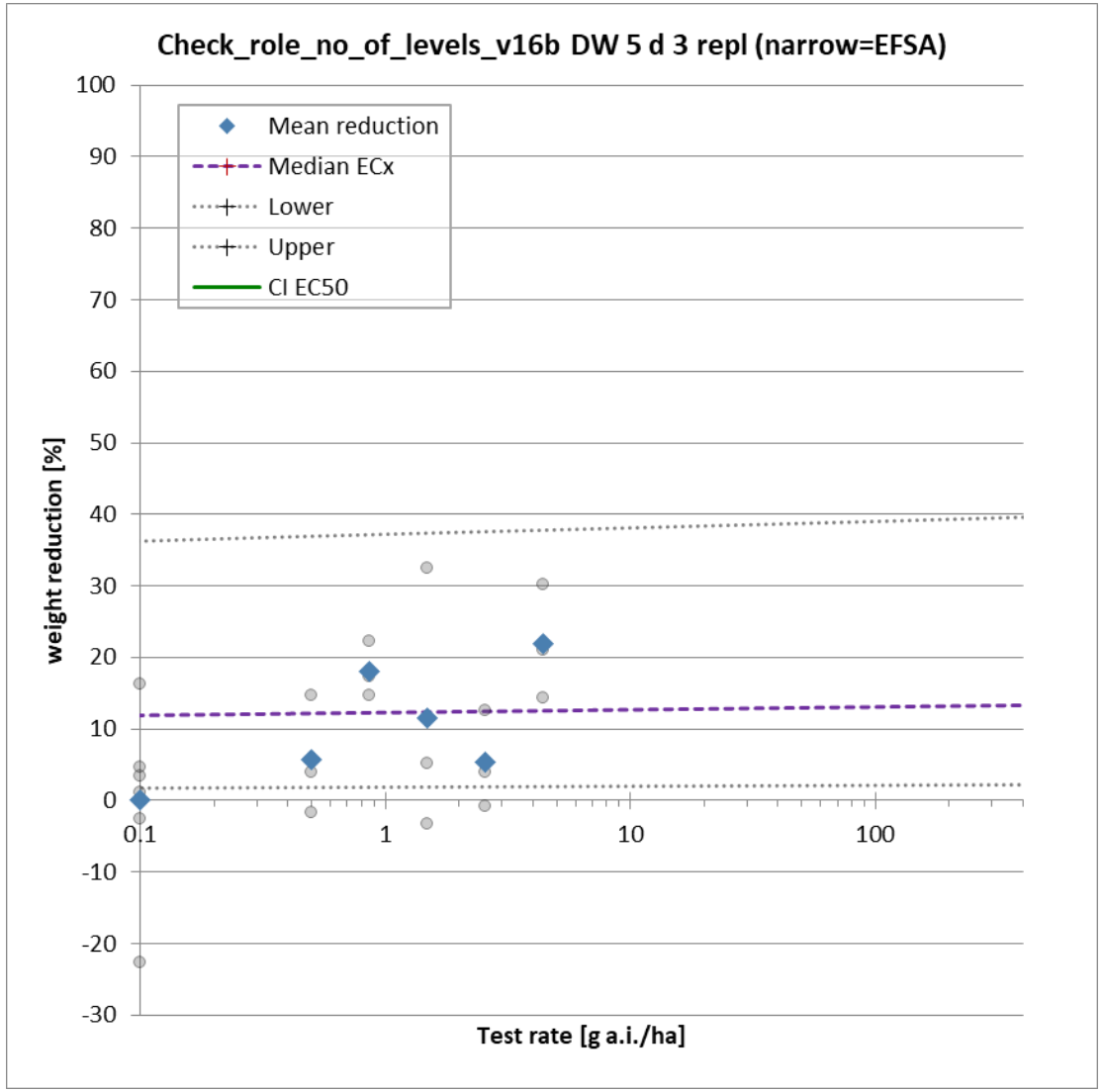
Confidence-intervals of ECx

- ECx around EC10 and EC20 narrow if only that narrow range tested, (6 dose levels, 18 treated systems) but not that much better
- Wide test range with 13 dose levels (39 treated systems) not distinctly better than wide test range with 5 dose levels (15 treated systems)
- With narrow range…



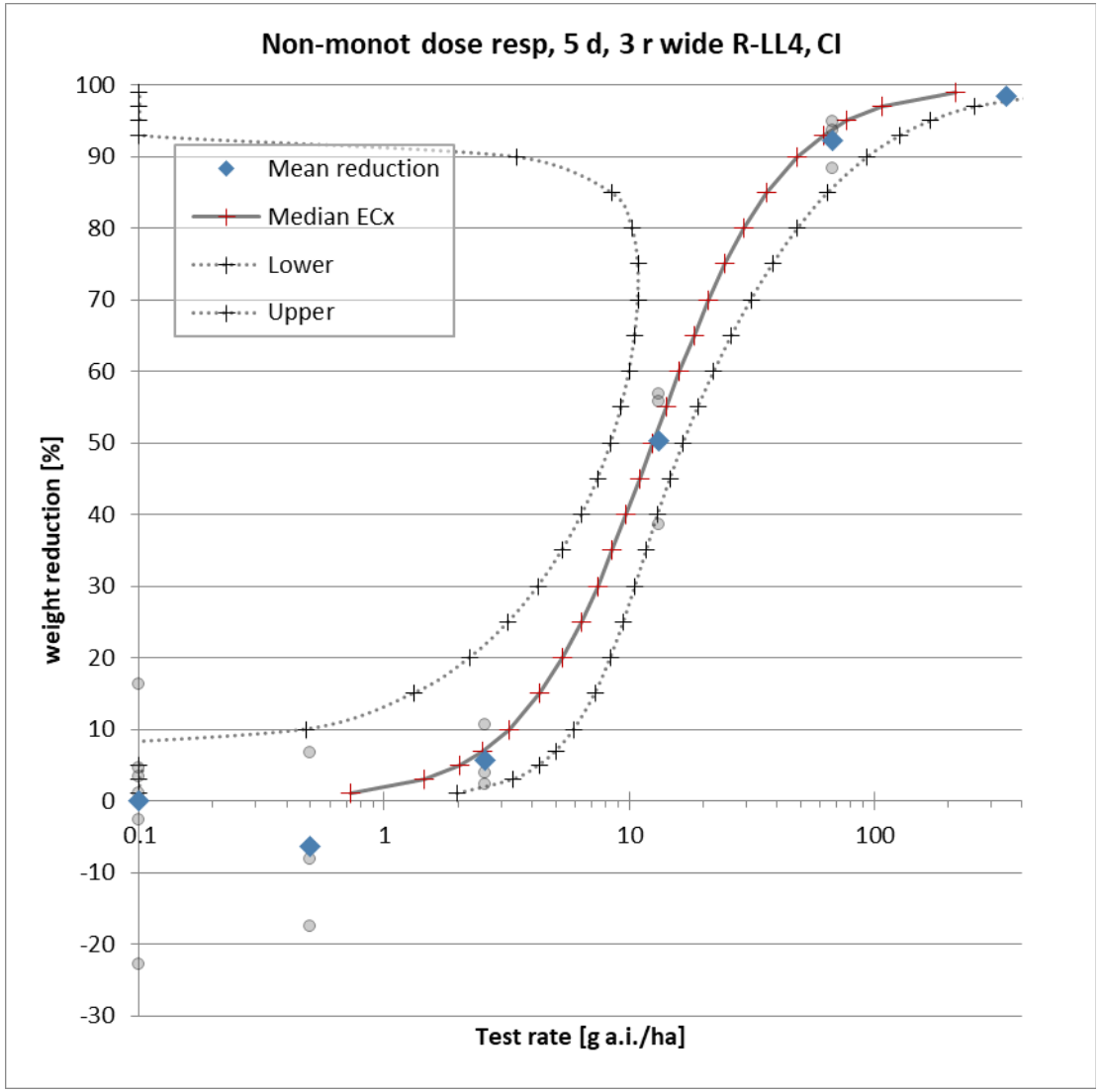
Confidence-intervals of ECx

- ECx around EC10 and EC20 narrow if only that narrow range tested, (6 dose levels, 18 treated systems) but not that much better
- Wide test range with 13 dose levels (39 treated systems) not distinctly better than wide test range with 5 dose levels (15 treated systems)
- With narrow range risk distinctly increased to obtain uninterpretable data “CI not determined for mathematical reasons” (e.g. negative estimates)



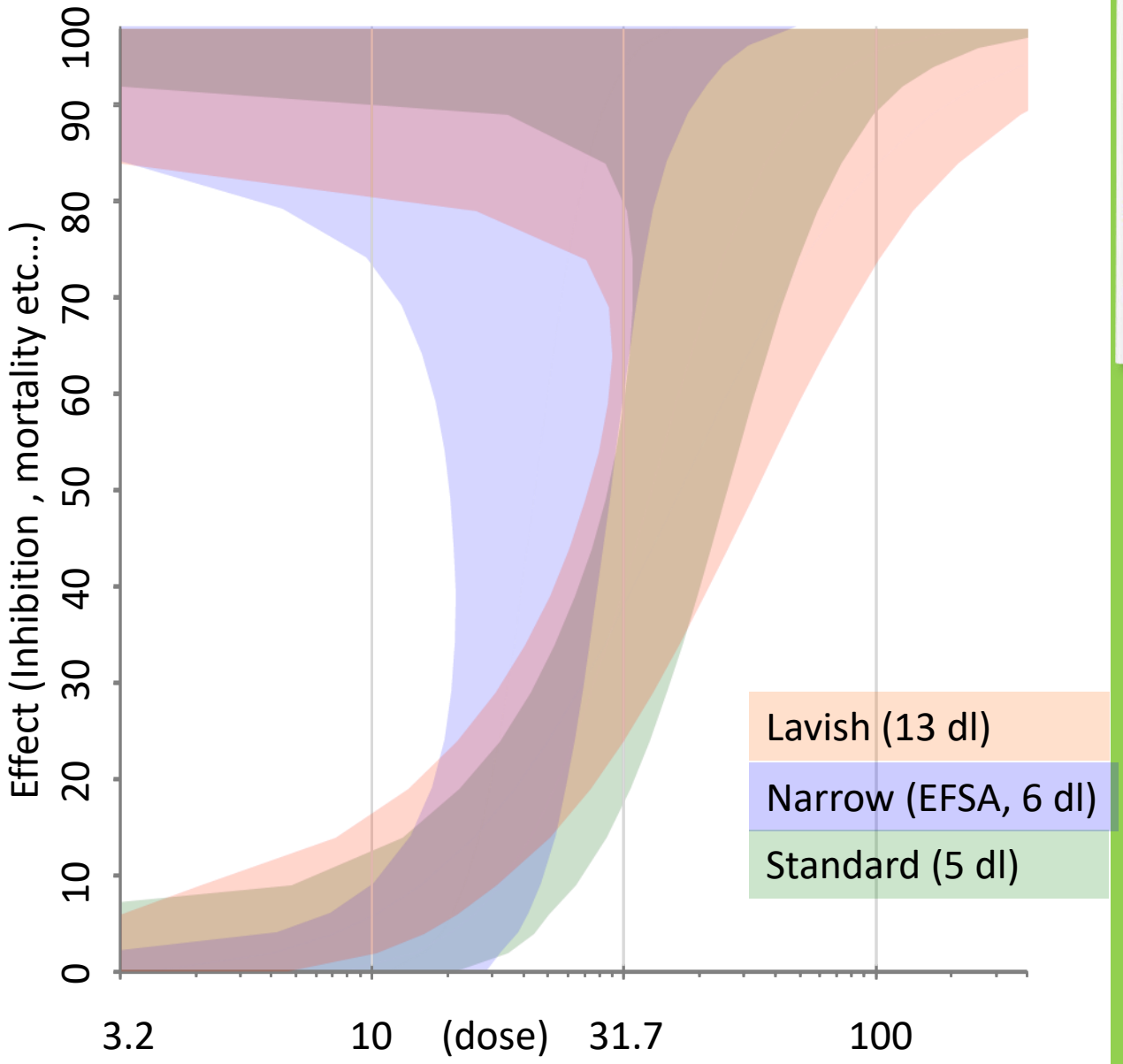
Confidence-intervals of ECx

- ECx around EC10 and EC20 narrow if only that narrow range tested, (6 dose levels, 18 treated systems) but not that much better
- Wide test range with 13 dose levels (39 treated systems) not distinctly better than wide test range with 5 dose levels (15 treated systems)
- With narrow range risk distinctly increased to obtain uninterpretable data “CI not determined for mathematical reasons” (e.g. negative estimates)
- Wide range with 5 dose levels almost as good as with 13 dose levels



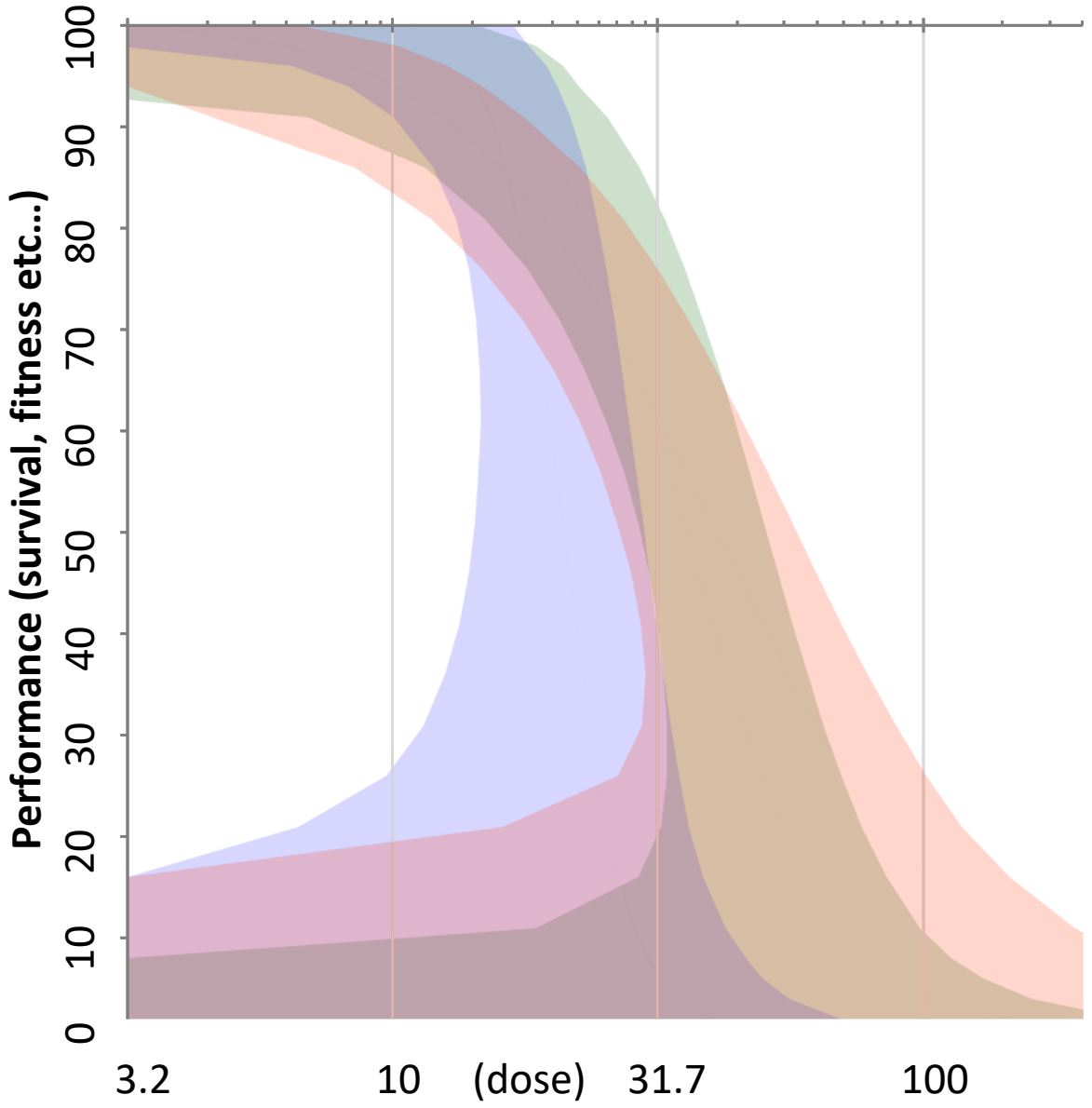
Confidence-intervals of ECx

- ECx around EC10 and EC20 narrow if only that narrow range tested, (6 dose levels, 18 treated systems) but not that much better
- Wide test range with 13 dose levels (39 treated systems) not distinctly better than wide test range with 5 dose levels (15 treated systems)
- With narrow range risk distinctly increased to obtain uninterpretable data “CI not determined for mathematical reasons” (e.g. negative estimates)
- Wide range with 5 dose levels almost as good as with 13 dose levels



Confidence-intervals of ECx

- ECx around EC10 and EC20 narrow if only that narrow range tested, (6 dose levels, 18 treated systems) but not that much better
- Wide test range with 13 dose levels (39 treated systems) not distinctly better than wide test range with 5 dose levels (15 treated systems)
- With narrow range risk distinctly increased to obtain uninterpretable data “CI not determined for mathematical reasons” (e.g. negative estimates)
- Wide range with 5 dose levels almost as good as with 13 dose levels



Confidence-intervals

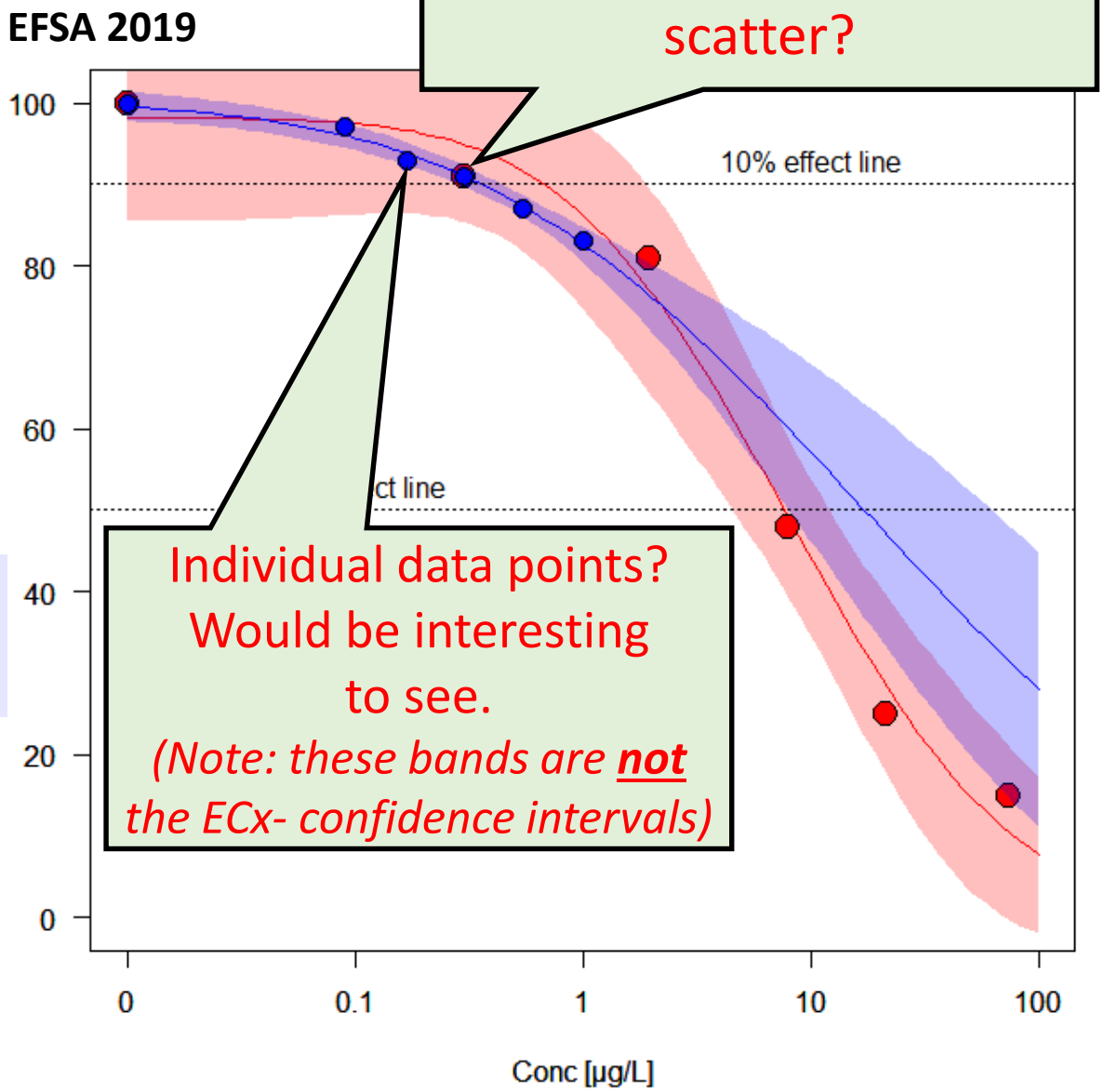
Outcome of pesticides peer review meeting on recurring issues in ecotoxicology

EFSA 2009, p 60, (above Fig. F8):
"However, blue points are clustered within a concentration range that yields between 0 and 20% effects. Red points are more spaced, and the highest concentration yields up to 75% effect.

Blue points can describe better the dose-response curve between 0% and 20% effect? (That much improvement? I wonder…)

Indeed confidence interval around EC10 is very narrow compared to the (Same scatter?) confidence provided by the red points.

However, at higher effect levels, the confidence provided by the blue points is rapidly decreasing and any ECX with X>20 is just a ventured guess which is likely not representing a reliable value (e.g. see the EC50 illustrated in Figure F8)" (Fully agree)



Conclusion 2:

Testing just a narrow range (0 – 20% effect)

- Only minor improvement of EC_{10} or EC_{20}
- Distinctly increased risk of more invalid runs

3: Cases with only 0% or 100% effect

EFSA-recommendation: repeat with additional intermediate test levels,
until at least two levels with partial effects

- Method: Deterministic survival data – varying individual mortality

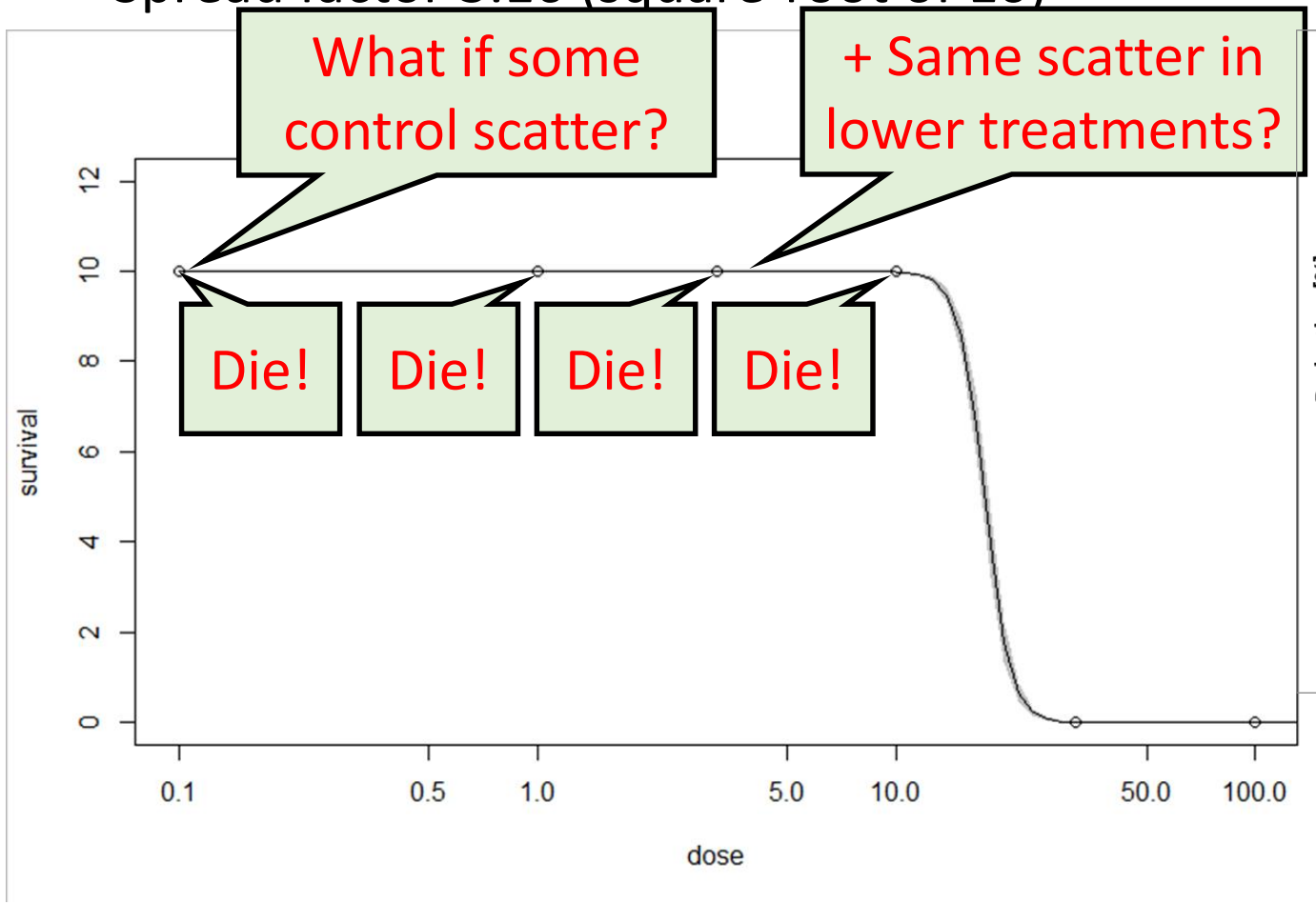
  - checking most extreme outcomes, thus
  - covering all possible outcomes of a given test design, e.g.

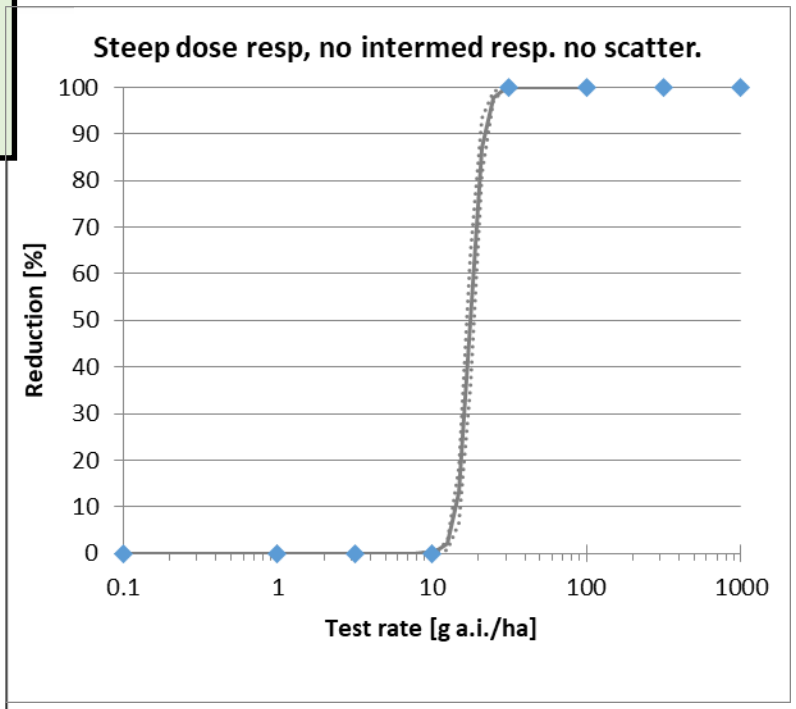
  6 control replicates,
  3 replicates per treatment level,
  10 specimens per replicate

Datasets with no partial effects

Standard test design: 3 replicates; 10 animals each (30 per dose level) 5 or more dose levels

Spread factor 3.16 (square-root of 10)



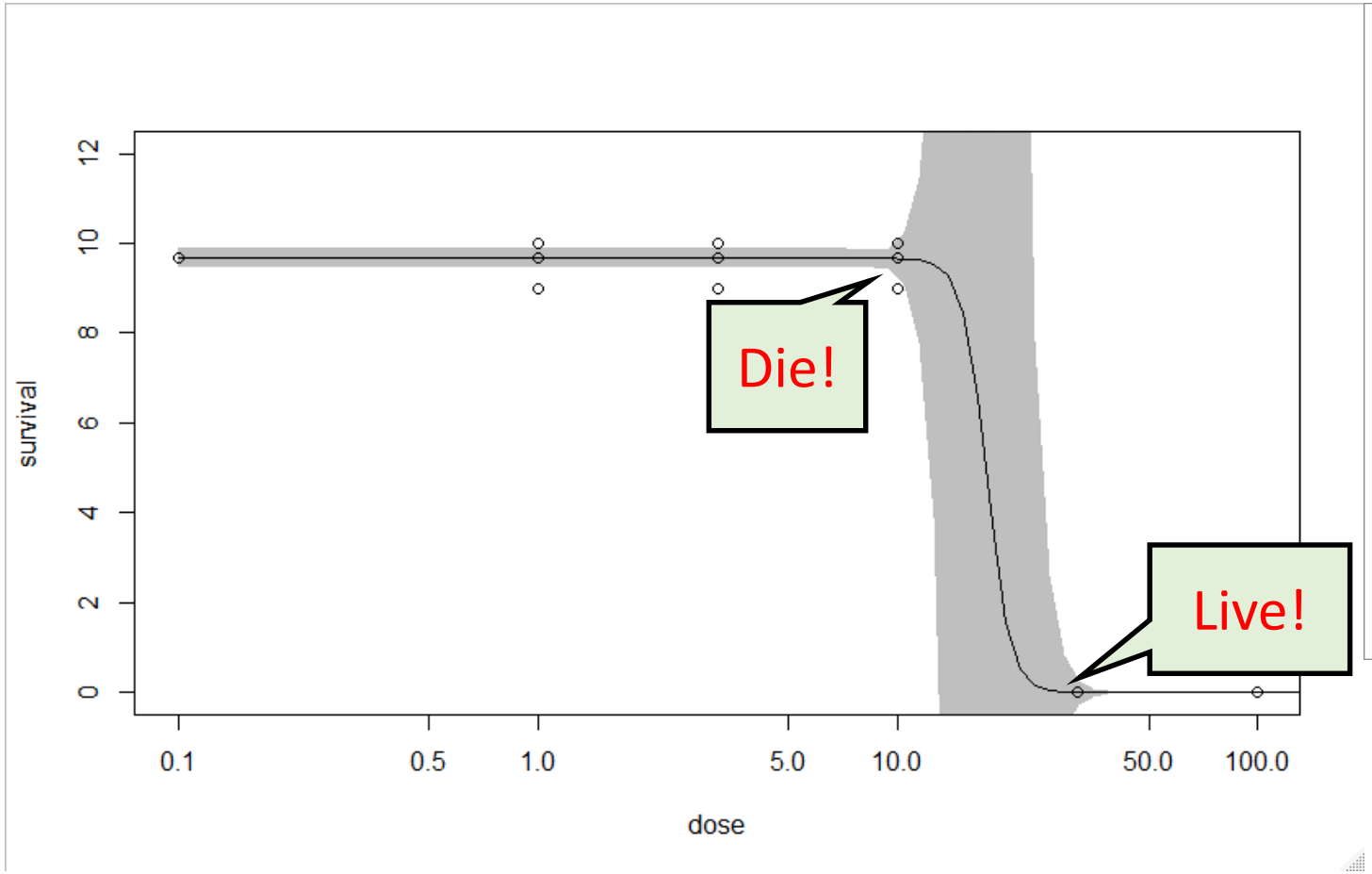


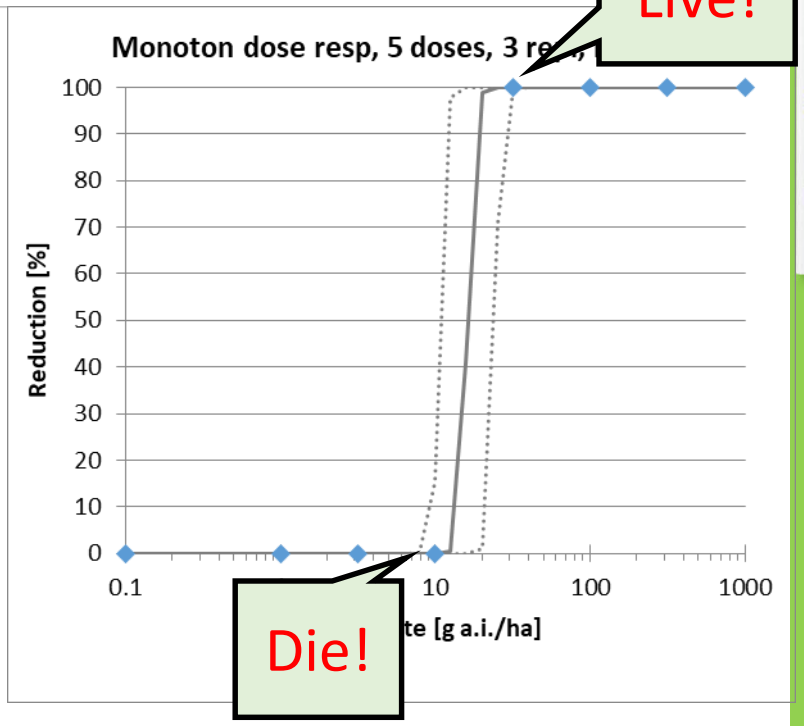
EC_{50} = 17.8 (geomean)

EC_{10} = 14.65 (r) (sd 0.186) 14.3 - 15.0 (factor 1.05 , NW_{CI} = 0.05)

Datasets with no partial effects

Spread factor 3.16 (square-root of 10)



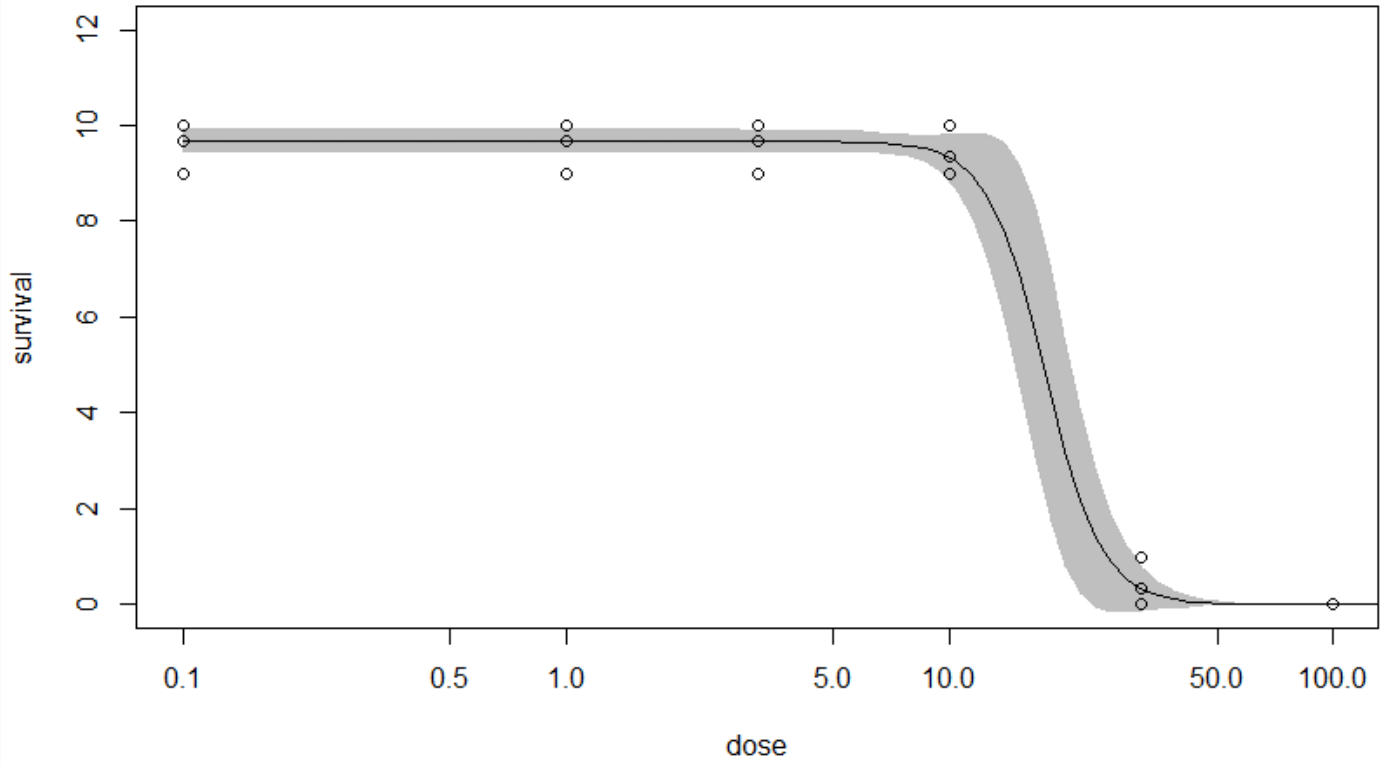


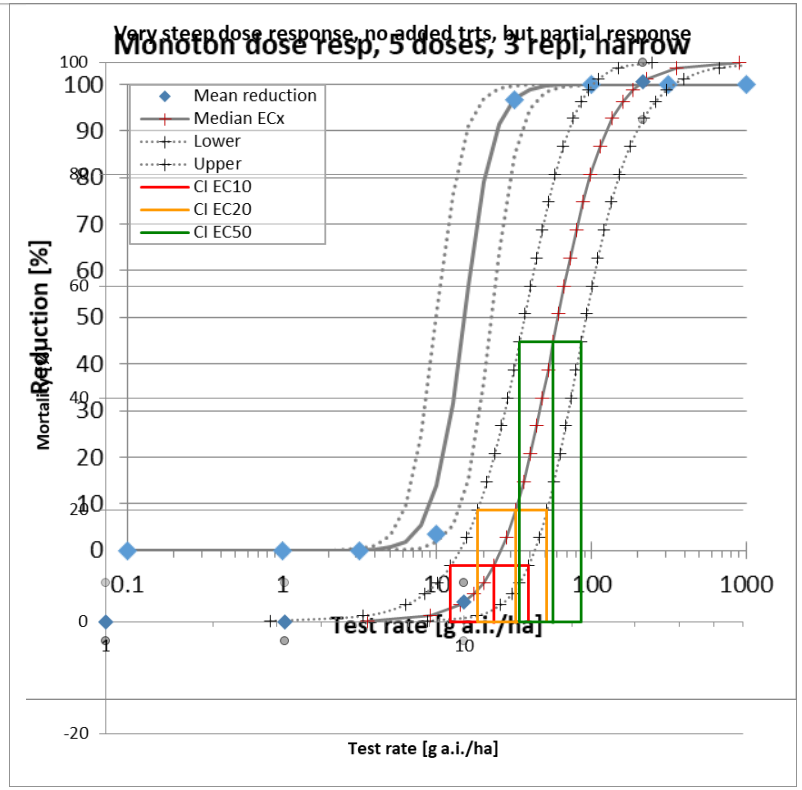
EC_{50} = 17.8 (geomean)

EC_{10} = 14.9 (r) (sd 23.7) -34.1 - 63.9 (factor ∞ !, NW_{CI} = 6.57)

Datasets with no partial effects

Spread factor 3.16 (square-root of 10) (with minimal partial effect: ok!)



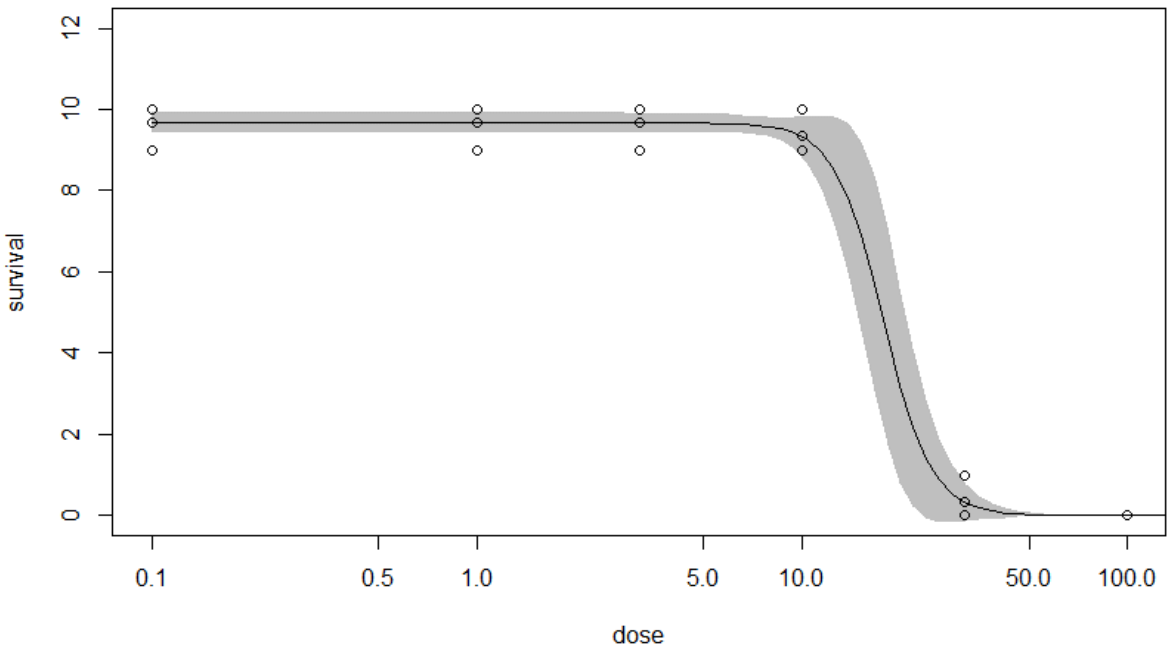


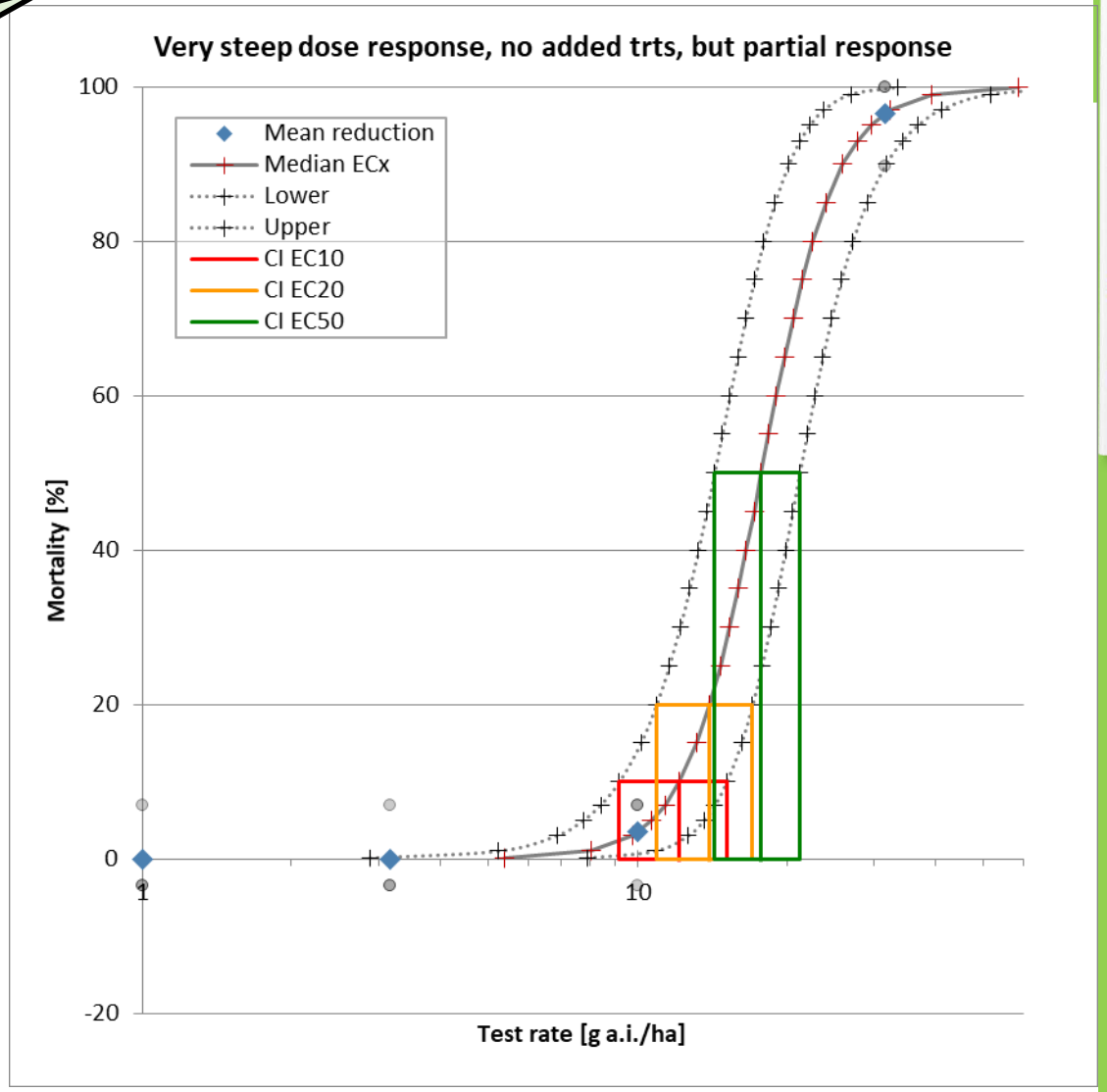
EC_{10} = 12.2 (sd 1.4) 9.2 - 15.1 (factor 1.65, NW_{CI} = 0.45)

Datasets with partial effects

Would be nice, but…

Spread factor 3.16 (square-root of 10)
(with minimal partial effect: ok!)



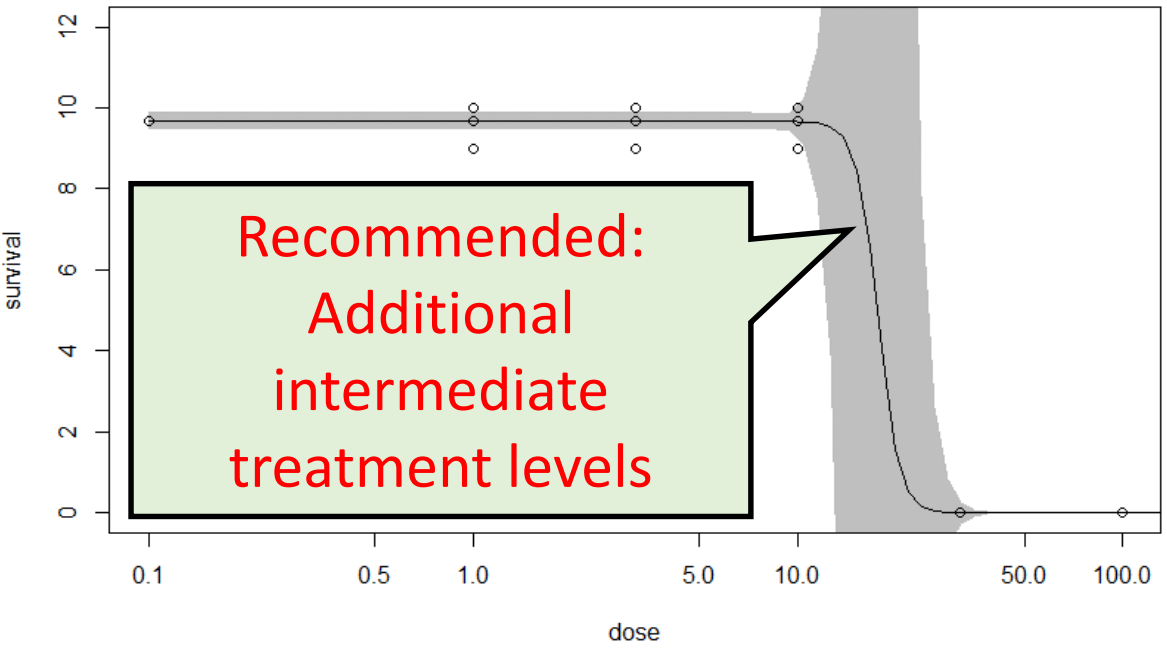


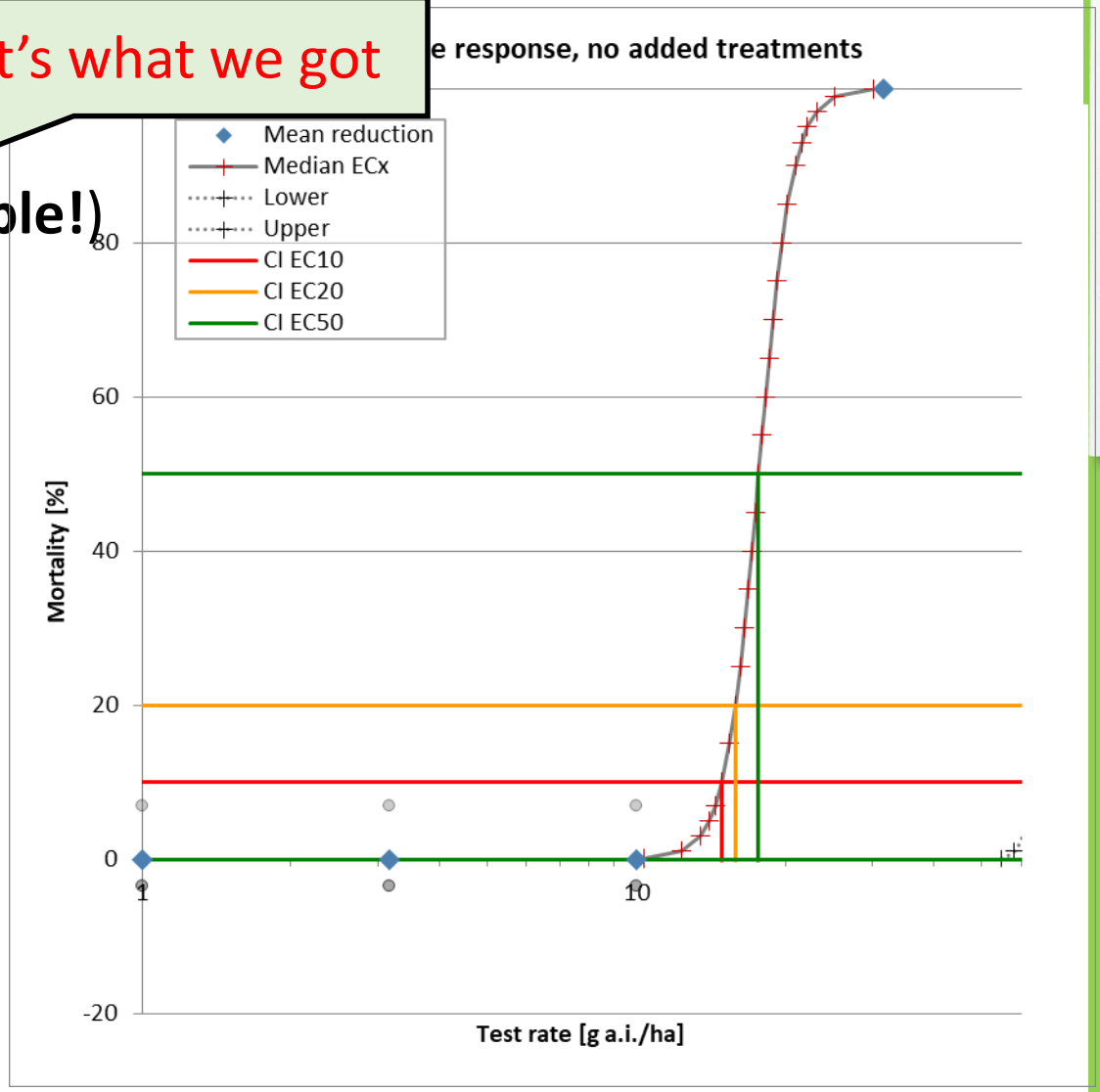
EC_{10} = 12.2 (sd 1.40) 9.2 - 15.1 (factor 1.65, NW_{CI} = 0.45)

Datasets with NO partial effects

…that's what we got

Spread factor 3.16 (10^(1/2))
(with minimal partial effect: **not acceptable!**)

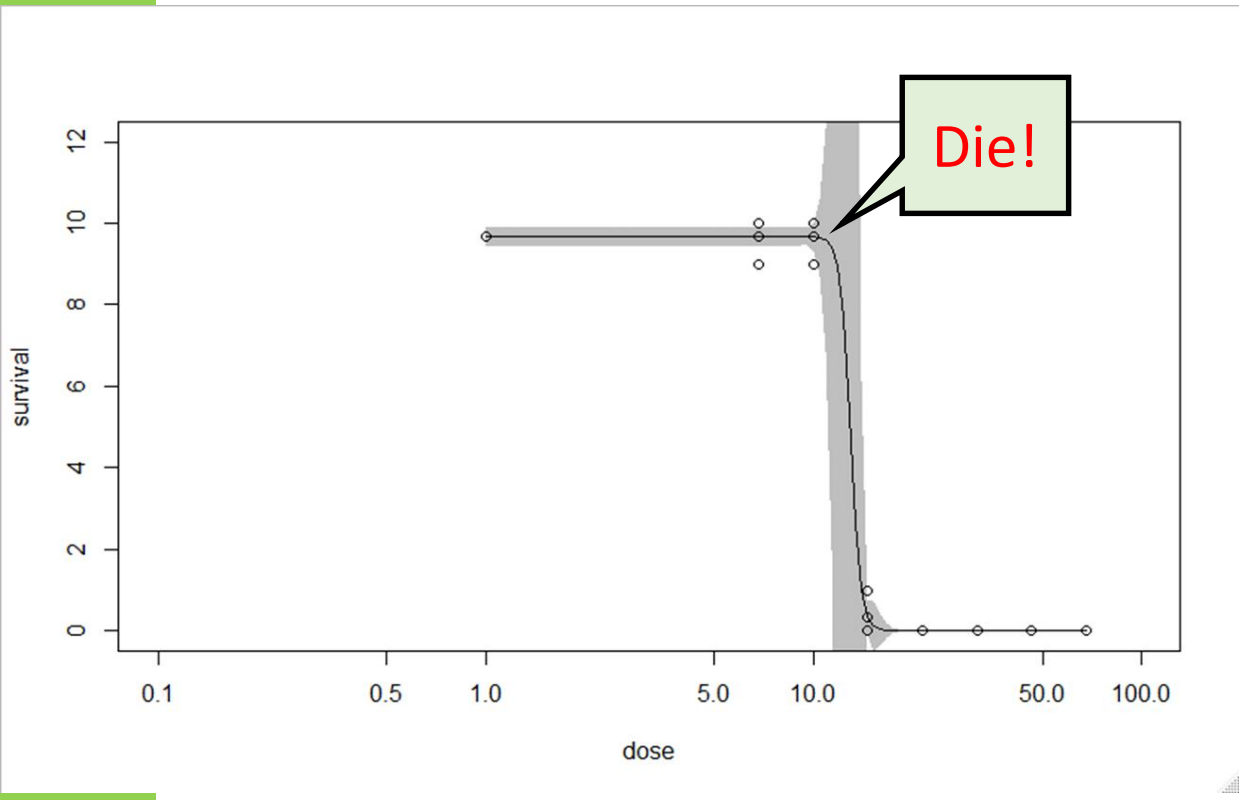


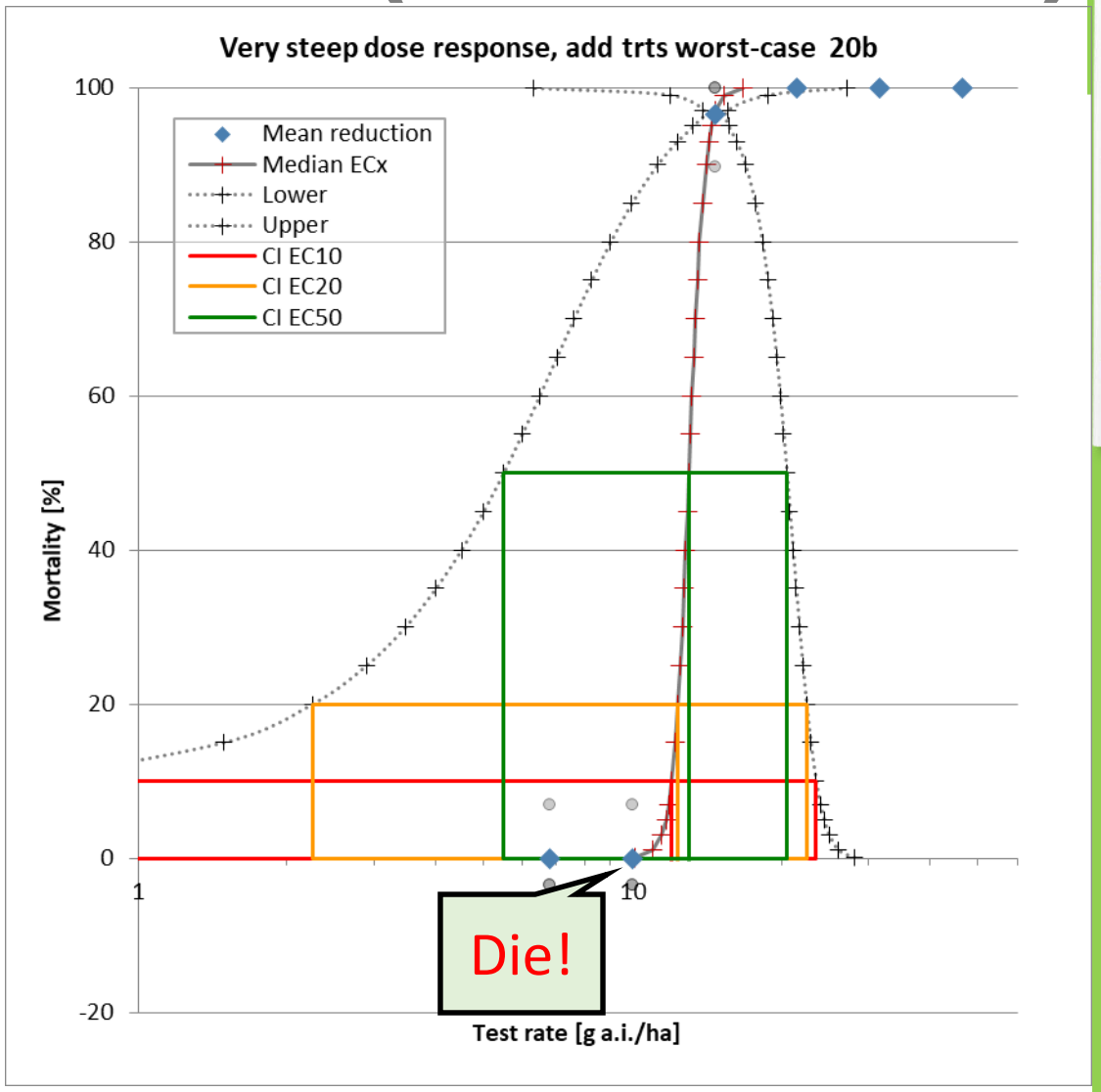


EC_{10} = 14.9 (r) (sd 23.7) -34.1 - 63.9 (factor ∞ !, NW_{CI} = 6.57)

Repeat with addit. levels (worst case)

Spread factor 1.47 (10^(1/6))
(with single partial effect: no luck)

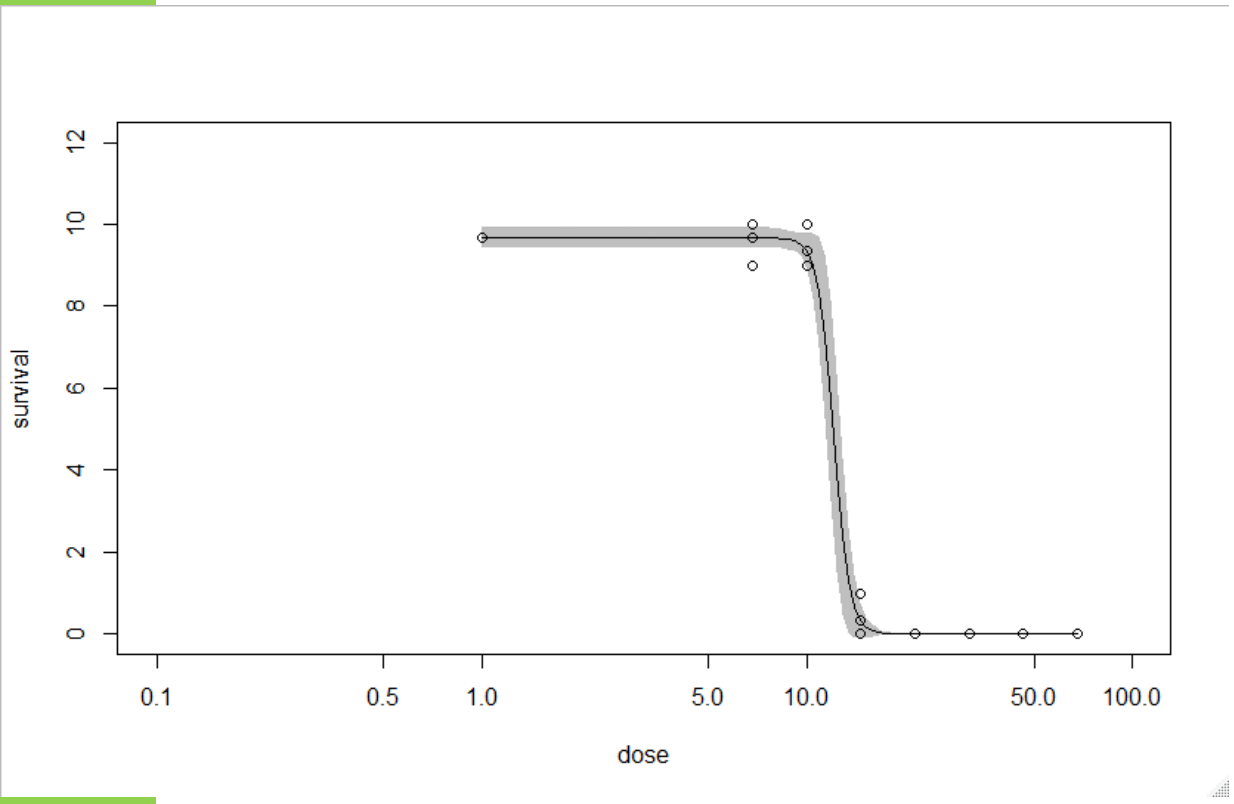


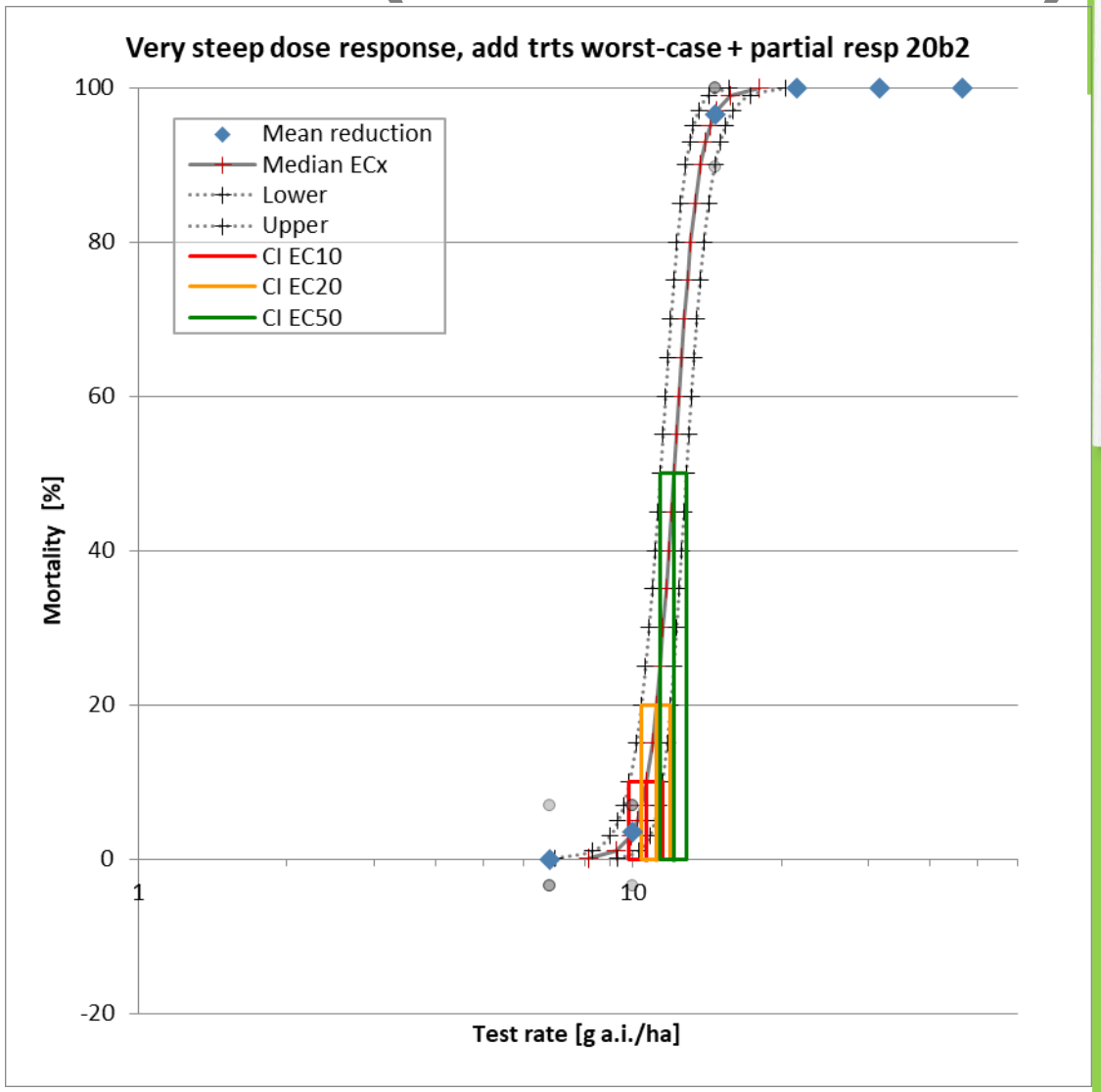


EC_{10} = 12.0 (sd 5.57) 0.50 - 23.5 (factor 46.7, NW_{CI} = 1.92)

Repeat with addit. levels (worst case)

Spread factor 1.47 (10^(1/6))
(with minimal partial effect: ok!)

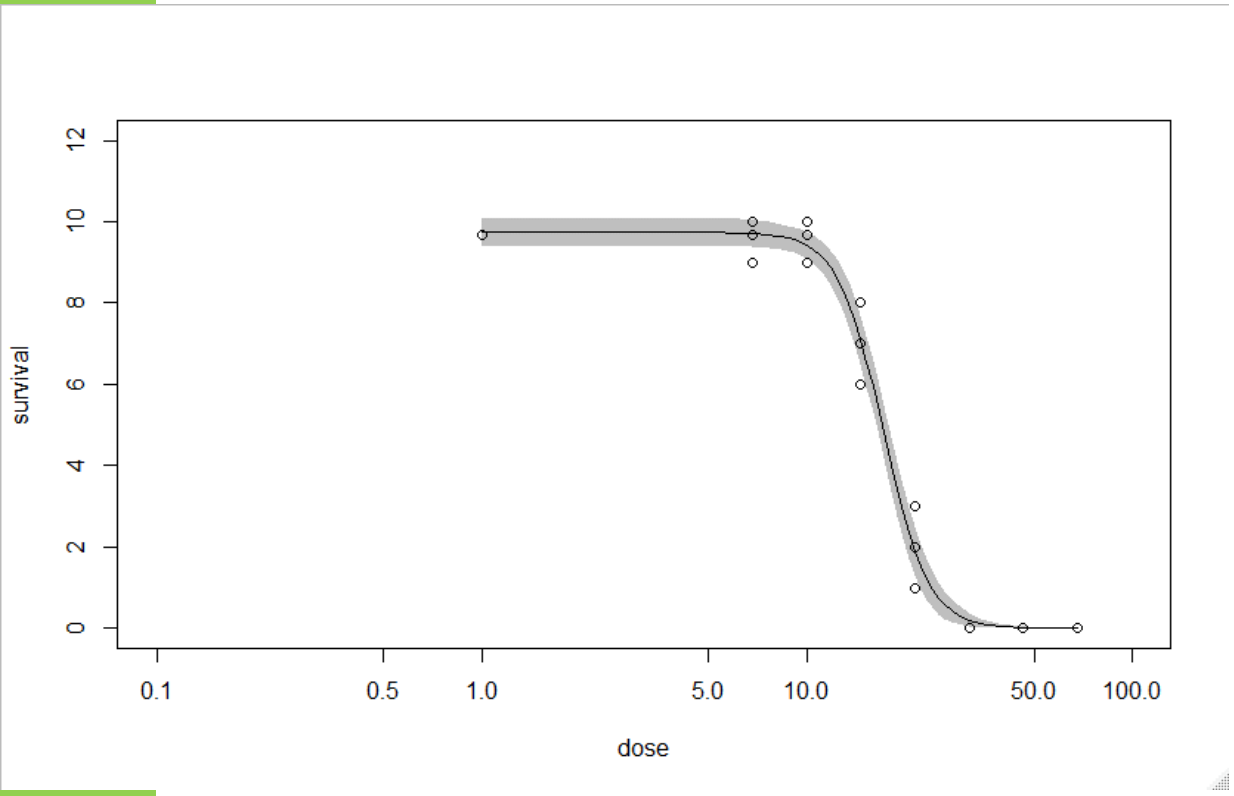


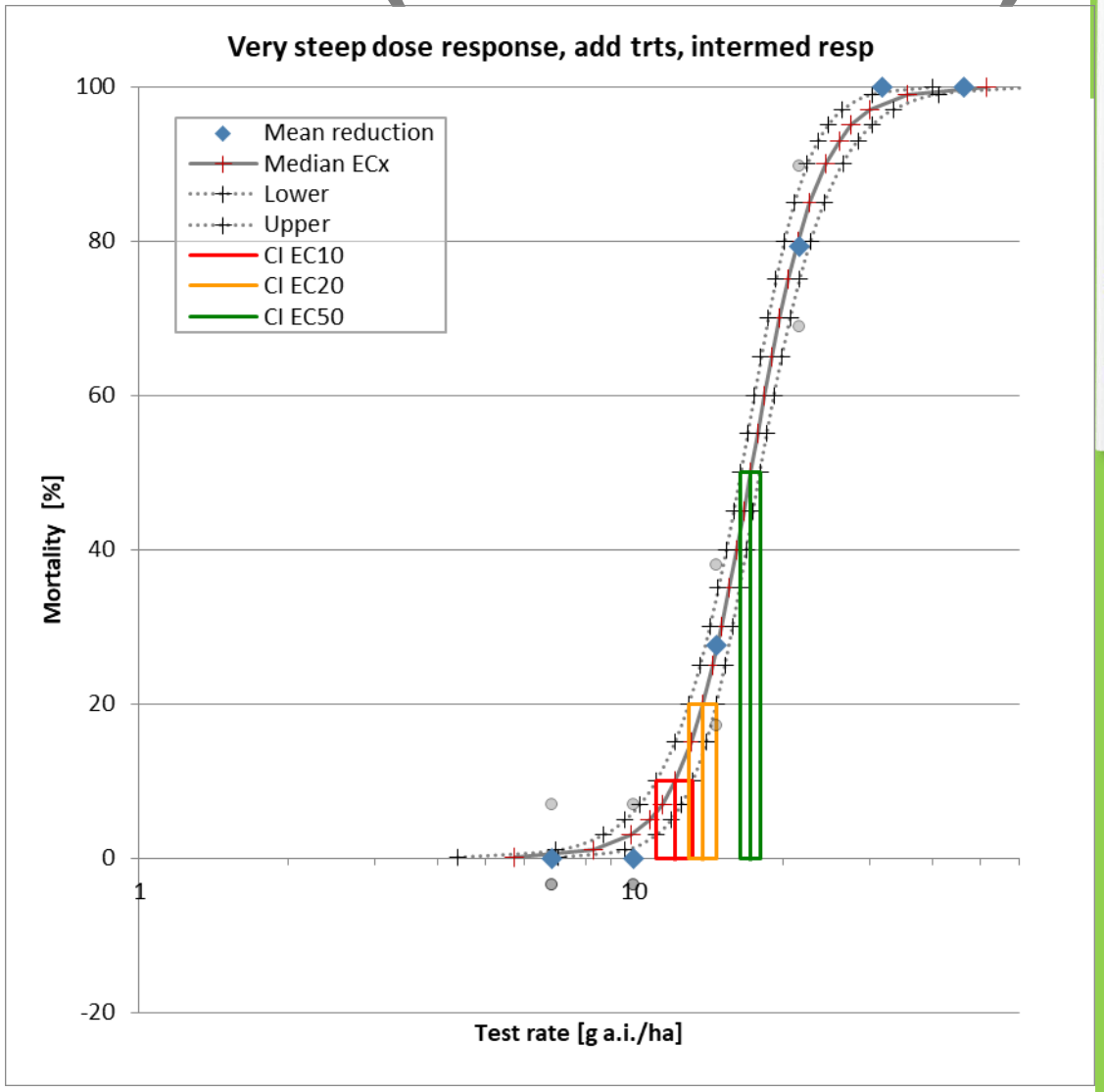


EC_{10} = 10.7 (sd 0.397) 9.85 - 11 (factor 1.17, NW_{CI} = 0.15)

Repeat with addit. levels (ideal case)

Spread factor 1.47 (10^(1/6))
(with ideal partial effect: ok!)

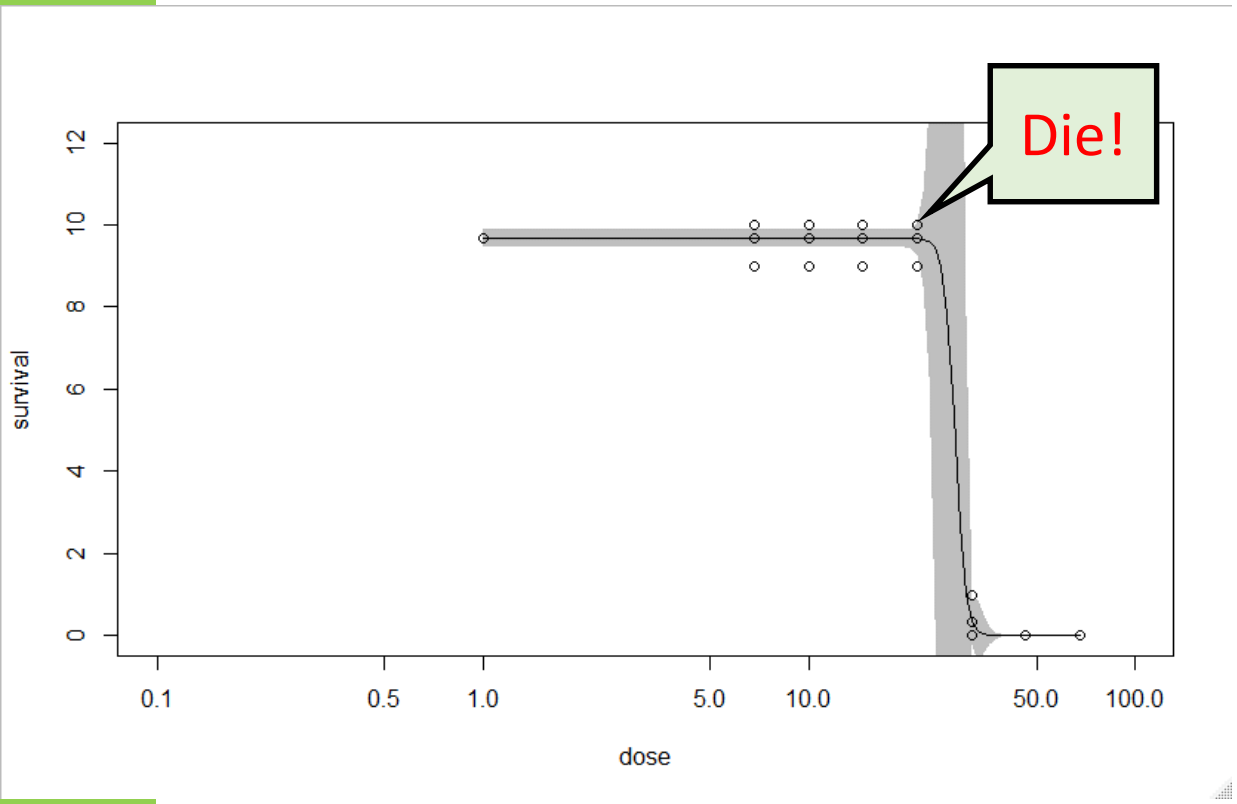


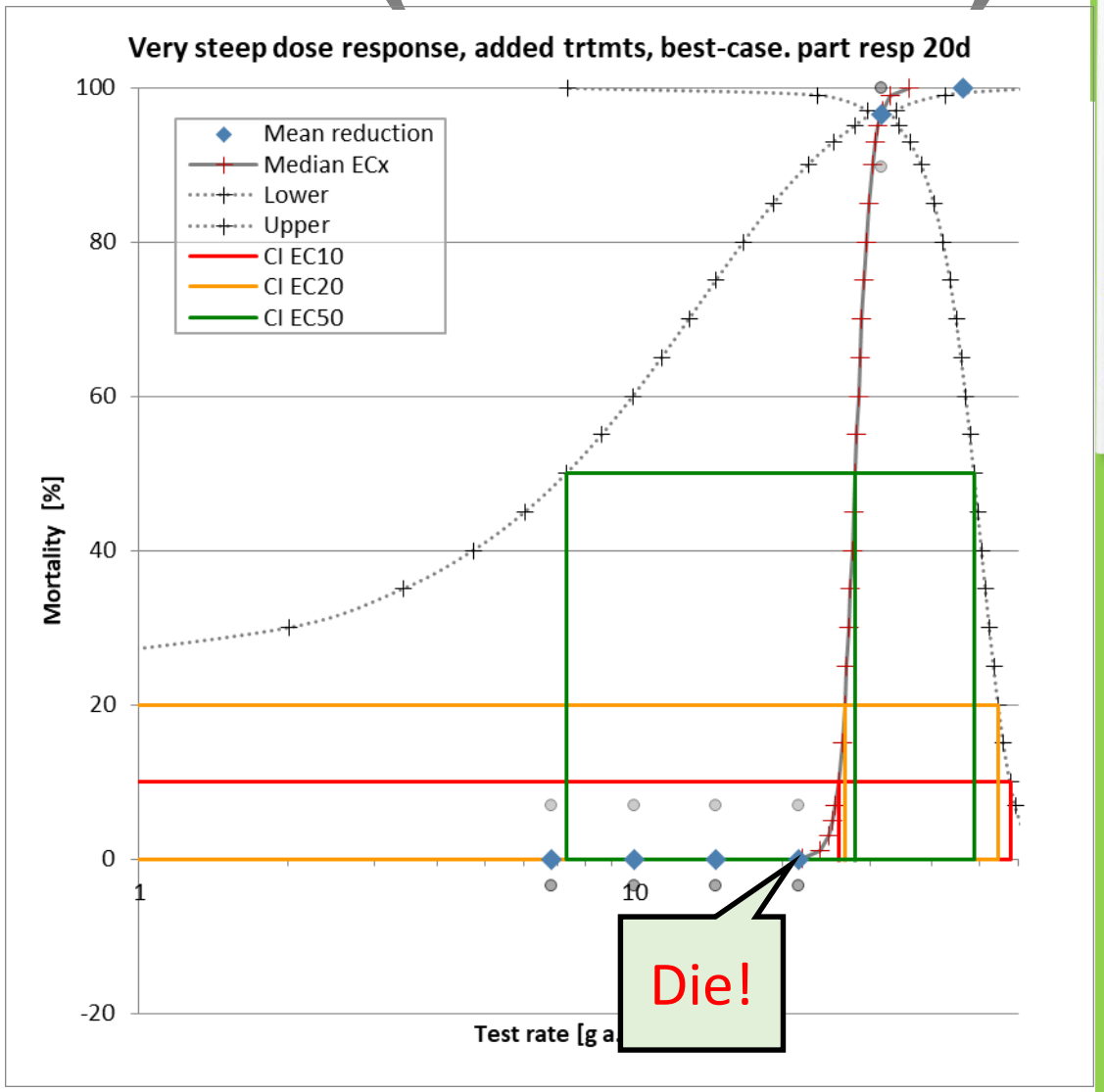


EC_{10} = 12.1 (sd 0.50) 11.1 - 13.2 (factor 1.18, NW_{CI} = 0.17)

Repeat with addit. levels (best case)

Spread factor 1.47 (10^(1/6))
(with some partial effect: not enough)

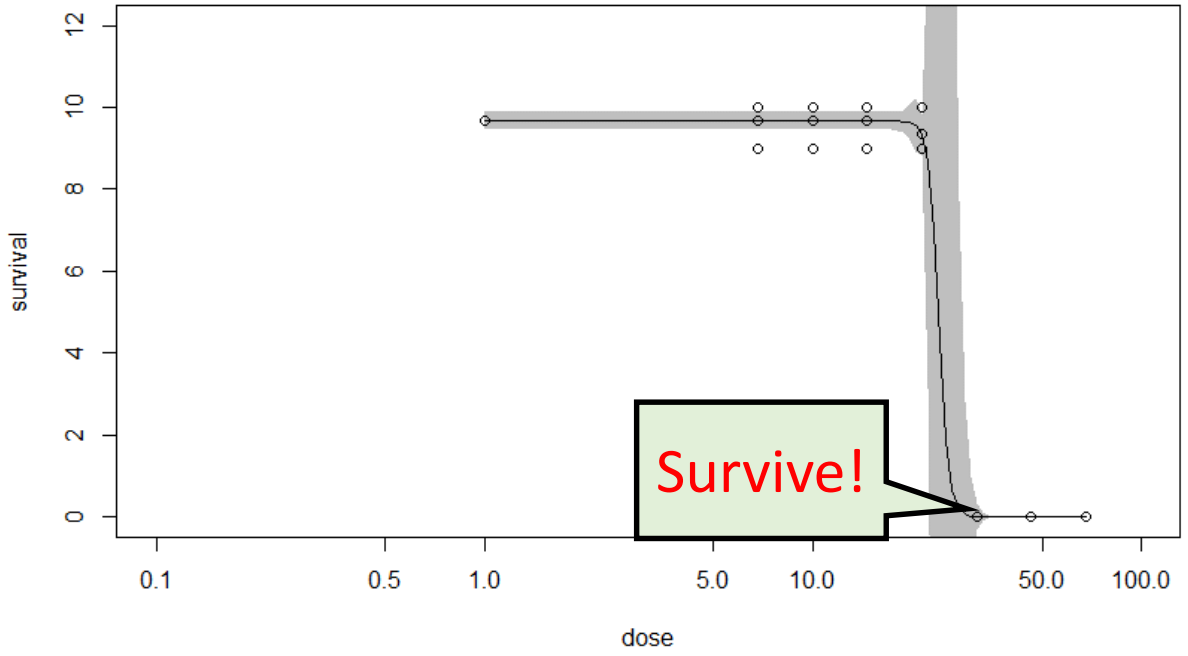


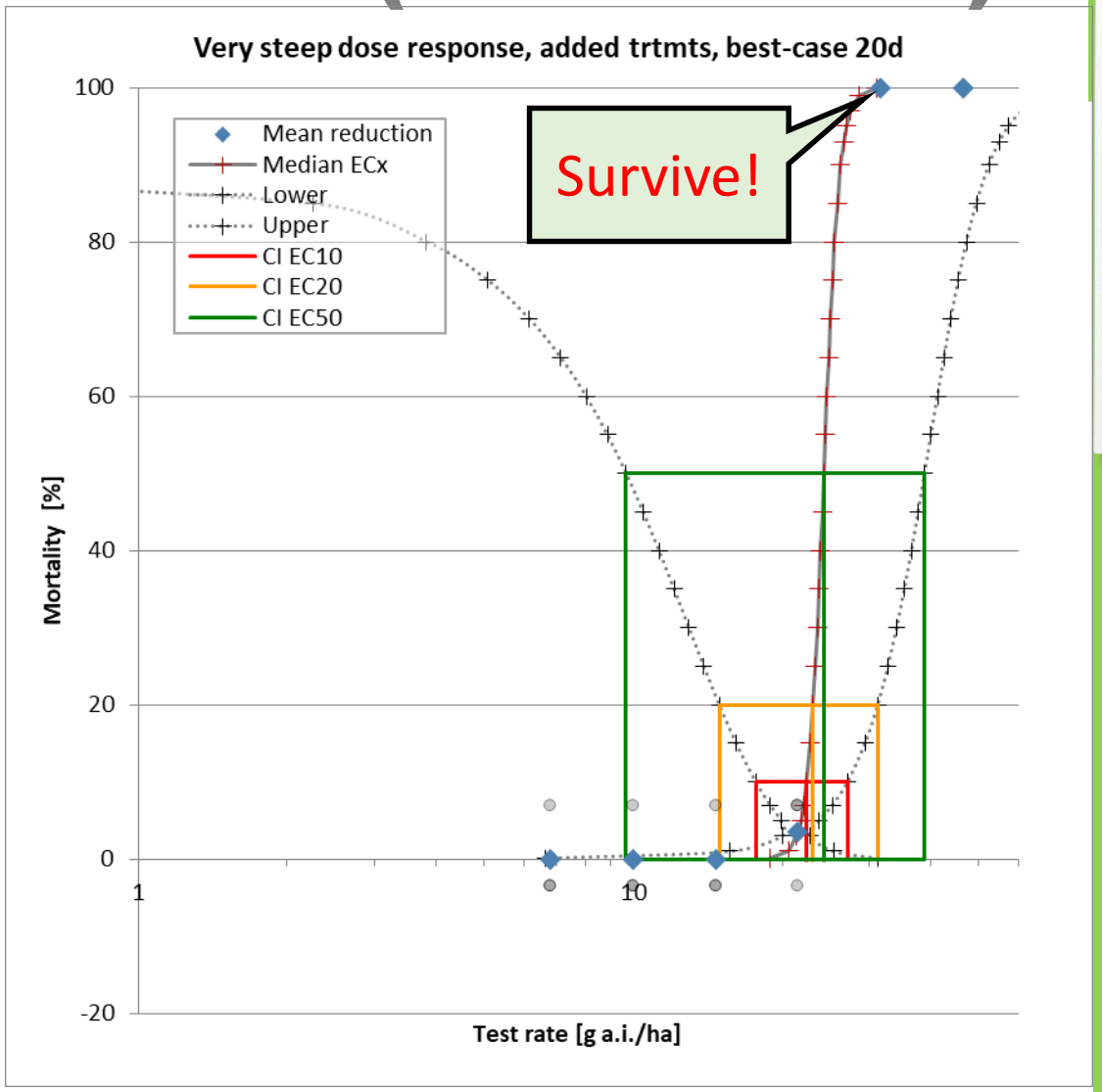


EC_{10} = 25.9 (sd 15.4) -5.8 - 57.7 (factor ∞ !, NW_{CI} = 2.45)

Repeat with addit. levels (best case)

Spread factor 1.47 (10^(1/6))
(with some partial effect: just enough)

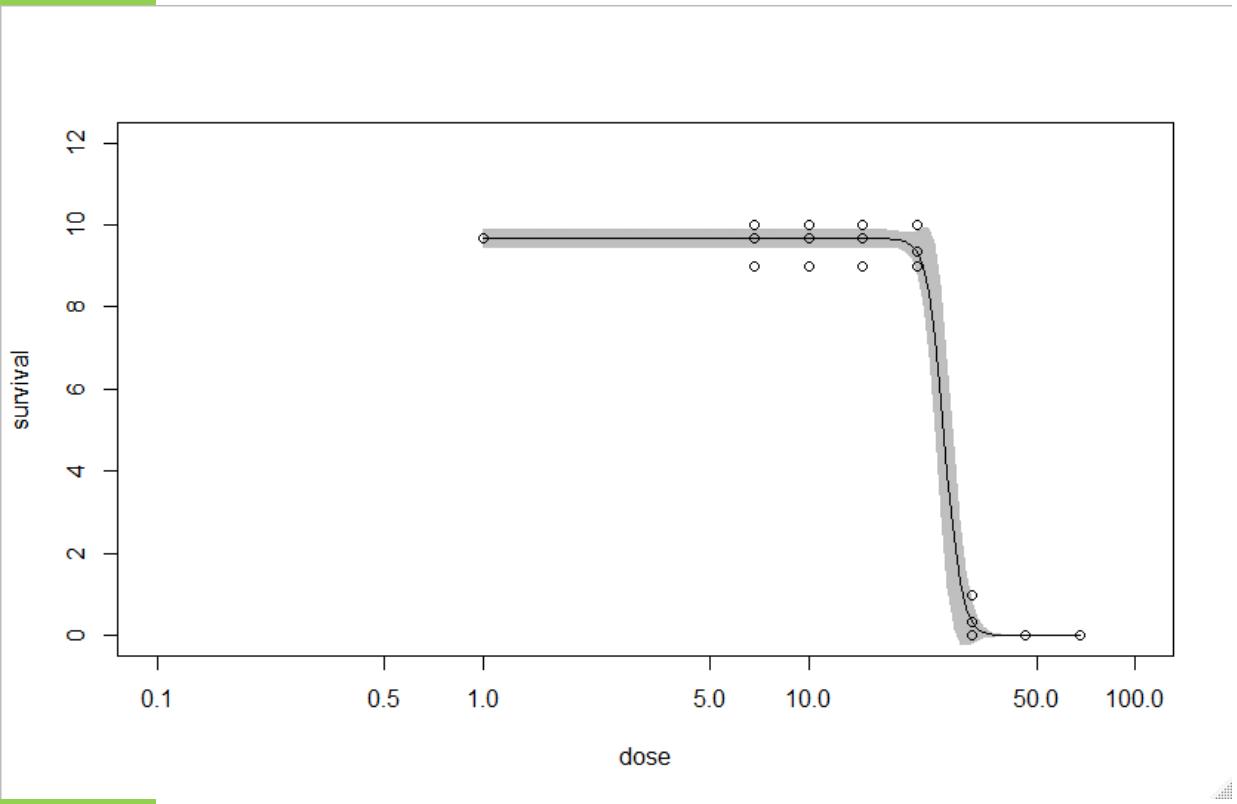


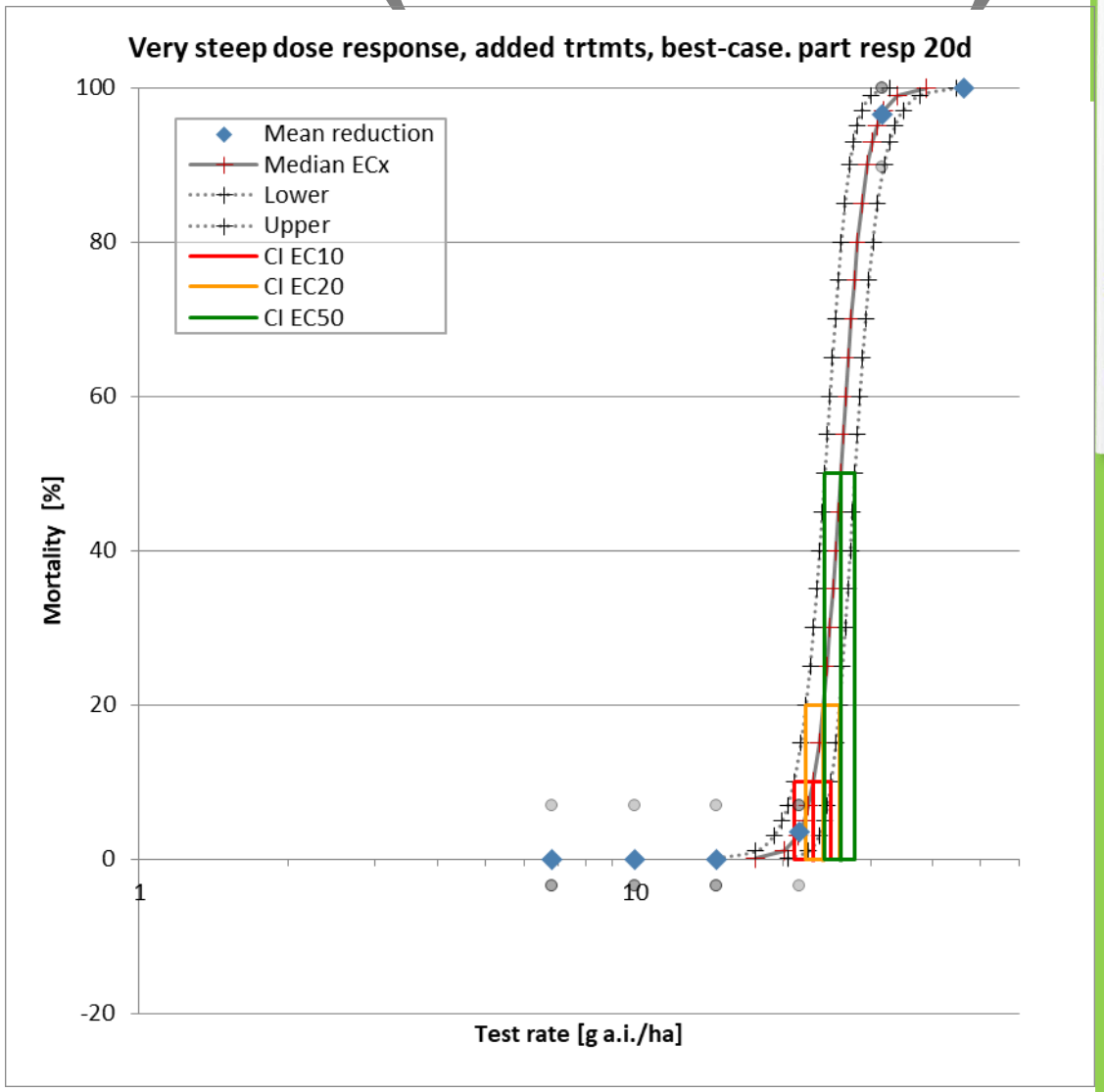


EC_{10} = 22.4 (sd 2.29) 17.7 - 27.2 (factor 1.50, NW_{CI} = 0.42)

Repeat with addit. levels (best case)

Spread factor 1.47 (10^(1/6))
(with enough partial effect: ok)





EC_{10} = 23.0 (sd 0.97) 21.0 - 25.0 (factor 1.19, NW_{CI} = 0.17)

More dose levels needed?

Deficiencies - if no levels with partial effect - only a mathematical problem:

Std. & geomean	EC_{50} = 17.8 (geomean), limits 10 – 31.6 (levels with 0 + 100% effect)			
Std. no scatter (LL.3)	EC_{10} = 14.7 (sd 1.86)	14.3	- 15.0	(factor 1.05, NW_{CI} = 0.05)
Std. (=!) w. r: (LL.3) ctr.s	EC_{10} = 14.9 (sd 23.7)	-34.1	- 63.9	(factor ∞ !, NW_{CI} = 6.57)
Std. LL.3 w. partial)	EC_{10} = 12.2 (sd 1.40)	9.2	- 15.1	(factor 1.65, NW_{CI} = 0.45)
Add dl, worst, no part.	EC_{10} = 12.0 (sd 5.57)	0.50	- 23.5	(factor 46.7, NW_{CI} = 1.92)
Add dl, worst + partial	EC_{10} = **10.7** (sd 0.40 = **min**	9.85	- 11.5	(factor 1.17, NW_{CI} = 0.15)
Add dl, ideal + partial	EC_{10} = 12.1 (sd 0.50)	11.1	- 13.2	(factor 1.18, NW_{CI} = 0.17)
Add dl, best, some part.	EC_{10} = 22.4 (sd 2.29)	17.7	- 27.2	(factor 1.50, NW_{CI} = 0.42)
Add dl, best, some part.	EC_{10} = **25.9** (sd 15.4 = **max**	-5.8	- 57.7	(factor ∞ !, NW_{CI} = 2.45)
Add dl, best + partial	EC_{10} = 23.0 (sd 0.97)	21.0	- 25.0	(factor 1.19, NW_{CI} = 0.17)
	So again: EC_{10} min & max:	10.7	- 25.9	
Std. & geomean:	**EC_{50} ≈ EC_{10} = 17.8** (geomean),	**10.0**	**- 31.6**	(levels with 0 + 100% effect)

…would have been perfectly fine, ***all*** possible *valid* outcomes (3 x 10 animals per dl) covered!

Conclusion 3:

If spreading factor is ≤ 3.2 and only 0% and 100% effect (very steep slope):
Geometric mean as EC_{50} (and even EC_{10}!) and adjacent levels as CI are covered.

If a dose-response is that steep, any repetition with further intermediate dose levels only proves the obvious.

A mathematically nice outcome is not the end in itself,

statistics are means to an end

Three individual Q & A:

1. Does ECx-estimate become more robust if more dose levels tested? (total no. of test systems unchanged, but fewer replicates)
   - Expectation: More dose levels = curve fit will be more robust

   ➢ Outcome: no benefit from more dose levels (if same total no. of test systems)

2. If EC_{10} or EC_{20} wanted (instead of NOEC+ LOEC), then test narrow range (only covering 0 to 20% effect)?
   - Expectation: Narrow test range = more certain EC_{10} or EC_{20}

   ➢ Outcome: little gain of precision, but sign. increased risk of invalid runs

3. If only 0% or 100% effect, does it help to repeat with additional test levels, aiming for at least two levels with partial effects
   - Expectation: Additional test levels = More precise results (EC_{10} or EC_{50})

   ➢ Outcome: Just a mathematical problem, little gain of precision; repetition = just proving the obvious

Take-home messages:

It is not expedient to overregulate

- Spacing (costly if flat dose-response),
- no. of dose levels (no benefit unless also total no. of test systems increased)
- Testing narrow ranges for EC_{10} or EC_{20} is dangerous; estimate only slightly more robust
- Steep dose responses are more certain as such (if spacing ≤3.2 & only 0% or 100% effect)
- Repetition with intermediate levels dispensable = just proving the obvious

Thank you.

Comments and questions?