



Miljøministeriet  
Miljøstyrelsen

# Benchmark dose

## Bedre udnyttelse af data i pesticidregistrering og -regulering

Bekæmpelsesmiddel-  
forskning nr. 198

April 2021

Udgiver: Miljøstyrelsen

Forfattere:

Signe M. Jensen<sup>1</sup>,  
Jens C. Streibig<sup>1</sup>,  
Nina Cedergreen<sup>1</sup>  
Christian Ritz<sup>2</sup>

<sup>1</sup>Institut for Plante- og Miljøvidenskab, Københavns Universitet

<sup>2</sup>Institut for Ernæring og idræt, Københavns Universitet

ISBN: 978-87-7038-300-4

Miljøstyrelsen offentliggør rapporter og indlæg vedrørende forsknings- og udviklingsprojekter inden for miljøsektoren, som er finansieret af Miljøstyrelsen. Det skal bemærkes, at en sådan offentliggørelse ikke nødvendigvis betyder, at det pågældende indlæg giver udtryk for Miljøstyrelsens synspunkter. Offentliggørelsen betyder imidlertid, at Miljøstyrelsen finder, at indholdet udgør et væsentligt indlæg i debatten omkring den danske miljøpolitik.

Må citeres med kildeangivelse

# Indhold

<b>1.</b>	<b>Sammendrag og konklusion</b>	<b>5</b>
<b>2.</b>	<b>Summary and conclusion</b>	<b>7</b>
<b>3.</b>	<b>Indledning</b>	<b>9</b>
<b>4.</b>	<b>Eksisterende litteratur på metoder omkring BMD</b>	<b>13</b>
4.1	Udledning/bestemmelse af BMD	13
4.1.1	Valg af definition af BMD for et på forhånd fastlagt BMR	13
4.1.2	Fastsættelse af BMR	16
4.1.3	Valg af dosis-respons-model	16
4.1.4	Estimering af BMDL	18
4.2	Andre betragtninger	20
4.2.1	Design af eksperimenter	20
4.2.2	Model valg	22
4.2.3	Model-gennemsnit	22
4.2.4	Hierarkiske eksperimenter og inklusion af kovariat information	24
4.2.5	Kombination af flere datasæt og integration af historiske data	25
4.2.6	Simultan inferens	26
4.2.7	Retningslinjer og regulativer	27
4.3	Opsummering	28
<b>5.</b>	<b><i>bmd</i> – en R-pakke til bestemmelse af BMD</b>	<b>29</b>
5.1	Baggrund	29
5.1.1	Eksisterende løsninger	29
5.1.2	Problemstilling	30
5.2	R-pakken	31
5.2.1	Nøglefunktionen <i>bmd()</i>	31
5.2.2	Definitioner af BMD	31
5.2.3	Bestemmelse af baggrundsniveauet	32
5.2.4	Beregning af BMDL	32
5.2.5	Bestemmelse af BMD ved model-gennemsnit	32
5.3	Opsummering	33
<b>6.</b>	<b>Eksempler på brug af BMD i økotoxikologi</b>	<b>34</b>
6.1	Binomiale data: Regnorme test på chloroacetamide	34
6.1.1	Problemstilling	34
6.1.2	Data	34
6.1.3	Analyse af data	34
6.2	Tælle data: Toksicitet test med kobber under varierende temperatur	35
6.2.1	Problemstilling	35
6.2.2	Data	36
6.2.3	Analyse af data	36
6.3	Kontinuerte data og model-gennemsnit: Fiske test	37
6.3.1	Problemstilling	37
6.3.2	Data	38

6.3.3	Analyse af data	38
6.4	Binomiale data i et hierarkisk design: En akut toksicitet test på $\alpha$ -cypermethrin	40
6.4.1	Problemstilling	40
6.4.2	Data	40
6.4.3	Analyse af data	40
6.5	Tid-til-event data: Immobilisering af daphnia som reaktion på $\alpha$ -cypermethrin	41
6.5.1	Problemstilling	41
6.5.2	Tid til immobilisering af daphnia	42
6.5.3	Analyse af data	42
6.6	SSD	44
6.6.1	Problemstilling	44
6.6.2	Data	45
6.6.3	Analyse af data	45
6.7	Opsummering	51
<b>7.</b>	<b>Sammenligning af forskellige metoder til bestemmelse af BMDL</b>	<b>52</b>
7.1	Opsætning af simulationer	52
7.2	Resultater	55
7.2.1	Effekten af antal gentagelser	55
7.2.2	Størrelsen på BMR	57
7.2.3	Variationen i data	59
7.2.4	Indflydelsen af den underliggende model	61
7.3	Opsummering	63
<b>8.</b>	<b>Diskussion</b>	<b>64</b>
<b>9.</b>	<b>Konklusion</b>	<b>66</b>
<b>10.</b>	<b>Litteratur</b>	<b>67</b>
<b>Bilag 1.</b>		<b>74</b>
Bilag 1.1	Forkortelser	74

# 1. Sammendrag og konklusion

Formålet med dette projekt har været at forbedre vidensgrundlaget for den aktuelle pesticidregulering ved at introducere benchmark dose (BMD) metoden for en mere datadrevet og objektiv vurdering med en dertil hørende potentiel reduktion i miljørisiko og påvirkning med pesticider. Introduktionen af BMD metoden skal ske ved at bygge bro mellem det anvendte aspekt og den teoretiske udvikling af BMD metoder og forsyne økotoksikologer, forskere og myndigheder med tilgængelige og praktiske værktøjer gennem implementering af metoderne i tilgængelige open source-softwaremoduler.

BMD metoder bruges til på baggrund af en dosis-respons model at finde doser som udgangspunkt for en farlighedsvurdering i risikovurderingsprocessen. BMD er defineret som den dosis, der resulterer i en lille på forhånd fastlagt "acceptabel" stigning i sandsynligheden for et uønsket event i forhold til baggrundssandsynligheden. I praksis bruges ofte den nedre grænse for konfidensintervallet for den estimerede BMD (BMDL) som udgangspunkt for grænseværdifastsettelse.

Selvom der i det sidste årti er sket fremskridt inden for BMD-metoderne, har vi i dette projekt identificeret en række områder, hvor der er presserende behov for yderligere arbejde. Blandt de tilbageværende udfordringer er implementering af nye statistiske metoder i brugervenlig software; ingen konsensus om, hvordan man kan udlede BMDL (hvilke konfidensintervaller, der skal bruges); meget konservativ justering, hvis flere BMR-værdier overvejes; ingen definition af BMD for event-time-data; og ingen generel modelleringsmetode til mere komplekse eksperimentelle design. Der mangler desuden grundige sammenligninger af forskellige metoder inden for BMD-metodologien.

Den udviklede R-pakke imødekommer nogle af de udfordringer relateret til bestemmelse af BMD og BMDL der ikke imødekommes af BMDS og PROAST, de for tiden bedste specialiserede software muligheder til BMD-bestemmelse. Derudover samler R-pakken mulighederne der fra de to software løsninger så man som bruger kan nøjes med at lave alle sine analyser ét sted. Da R-pakken *drc* i forvejen er et populært valg til dosis-respons analyser bliver den nye R-pakke en simpel udbygning af data analysen i et og samme software.

R-pakken *bmd* udbygger mulighederne og adskiller sig blandt andet fra de eksisterende specialiserede software programmer ved at kunne bestemme BMD for flere typer respons data, herunder tælle data og tid-til-event data, til data typer der tit optræder inden for økotoksikologi. R-pakken muliggør bestemmelse af BMD for stort set alle dosis-respons-funktioner, der er til rådighed i BMDS og PROAST. Derudover, giver den nye R-pakke mulighed for at bruge fraktionelle-polynomier, der er særligt velegnet til model-gennemsnit, og flere hormesis modeller, samt semi-parametriske alternativer. I den udviklede R-pakke er der desuden mulighed for at beregne BMDL baseret på sandwich-variens-covarians estimater, der giver robuste standardfejl og dermed gør det muligt at tage højde for model misspecifikationer relateret til fordelingsantagelserne. BMDL kan bestemmes ved flere forskellige former for bootstrap og der er flere muligheder for at bestemme BMD og ikke mindst BMDL gennem model-gennemsnit.

Fleksibiliteten af BMD-metoden til at håndtere forskellige analytiske problemstillinger relateret til data fra økotoksikologien er her illustreret gennem seks data eksempler. Dataeksemplerne dækker mange tænkelige eksempler fra økotoksikologien herunder det mest almindelige eksempel med en enkelt dosis-respons kurve for binomiale, men også tælle- og kontinuerte data. For tælle data bliver det mere specifikt illustreret hvordan man kan håndtere forskellige udfordringer relateret til reproduktionsdata. Desuden gennemgås bestemmelse af BMD og BMDL

ved modelgennemsnit samt hvordan data fra mere komplekse designs kan bruges til at bestemme BMD og BMDL. Endvidere berører et enkelt eksempel tid-til-event data, en data type der hidtil har været overset i BMD sammenhæng, mens det sidste eksempel illustrerer at SSD-analyser kan baseres på BMDL frem for ED<sub>50</sub> (eller ED<sub>x</sub>, NOAEL eller LOAEL). Det giver mere robuste hazard koncentrationer, da der i højere grad end tidligere tages højde for usikkerheden.

Baseret på et større simulationsstudie af de forskellige metoder til at bestemme BMDL fandt vi at invers regression ikke fungerer optimalt til BMDL estimation når der kigges på helt små ændringer i respons. Delta metoden leverer til gengæld rimelige dækningsgrader, særligt for "extra" definitionen hvor den øvre grænse udnyttes i definitionen. Til gengæld viste bootstrap sig at klare sig overraskende godt i stort set alle scenarier. Helt generelt viste simulationerne, at det er klogt at have så høje doser inkluderet at den øvre grænse af dosis-respons-sammenhængen kan fastsættes. Særligt hvis en sådan bruges i definitionen af BMD.

#### Konklusion:

BMD metoden har stort potentiale indenfor økotoksikologi, herunder som metode til at vurdere miljøeffekter i pesticidreguleringen. Der er udviklet en frit tilgængelig R-pakke der imødekommer flere af de problemstillinger der blev fundet ved en gennemgang af litteraturen indenfor BMD metoderne.

Rapporten dokumenterer pakken der blandt andet giver mulighed for at finde BMDL ved bootstrap-metoden der i et simulationsstudie resulterede konsistent gode resultater (dækningsgrader meget tæt på det nominelle niveau). Projektet har desuden vist at BMD-metoden er fleksibel nok til at håndtere analyse-mæssige problemstillinger i økotoksikologien, også for andre typer af data end der hvor metoden allerede anvendes på i dag. Dette er illustreret gennem typiske data eksempler fra økotoksikologien. En nyskabelse som kan forbedre tolkningen af de klassiske SSD analyser er, at man nu kan benytte BMDL i stedet for ED<sub>50</sub> eller NOAEL eller LOAEL.

## 2. Summary and conclusion

The objective this project has been to improve the knowledge base underlying the current pesticide regulation by introducing the benchmark dose (BMD) methodology for a more data driven and objective assessment with a corresponding potential reduction in environmental risk and impact of pesticides. The introduction of the BMD method is to be done by bridging the applied aspect and the theoretical development of BMD methods and providing ecotoxicologists, researchers and authorities with accessible and practical tools through the implementation of the methods in available open source software modules.

The BMD methodology is used to find regression-based doses as a starting point for a hazard assessment in the risk assessment process. BMD is defined as the dose that results in a small, predetermined "acceptable" increase in the probability of an adverse event relative to the background probability. In practice, the lower limit of the confidence interval for the estimated BMD (BMDL) is often used as the starting point for setting reference values.

Even though progress has been made in the BMD methodology over the last decade, we have, in this project, identified a number of areas where there is an urgent need for further work. Among the remaining challenges are the implementation of new statistical methods in user-friendly software; no consensus on how to derive BMDL (what type of confidence intervals to use); very conservative adjustment if multiple BMR values are considered; no definition of BMD for event-time data; and no general modeling approach for more complex experimental designs. Furthermore, in general more studies providing thorough comparisons of the different approaches is warranted.

The developed R package meets some of the challenges related to BMD and BMDL estimation that cannot be handled by the BMDS and PROAST, the two currently best, specialized software options for BMD estimation. In addition, the R package combined the opportunities from these two software solutions so that the user can do all analyses using one software. As the R-package *drc* is already a popular choice for dose-response analyses, the new R-package becomes a simple extension of the data analysis in the same software.

The R-package *bmd* extends the possibilities and differs from the existing specialized software programs by, among other things, being able to determine BMD for several types of response data, including count data and time-to-event data, two types of data that often occur in ecotoxicology. The R package enables estimation of BMD for practically all dose-response functions available in BMDS and PROAST. In addition, the new R package allows the use of fractional polynomials, particularly suitable for model averages, and several hormesis models, as well as semi-parametric alternatives. In the developed R package, it is also possible to calculate BMDL based on sandwich variance-covariance estimates that provide robust standard errors and thus allow for model misspecifications related to the distributional assumptions. Several different types of bootstrap can be used for estimating BMDL and several options for estimating BMD and, in particular, BMDL through model averaging are available.

The flexibility of the BMD method to handle various analytical issues related to data from ecotoxicology is illustrated through six data examples. The data examples cover many conceivable examples from the most common case with a single dose-response curve for binomial data, but also examples with count and continuous data. For count data, it is illustrated how to deal with various challenges related to reproduction data. In addition, determination of BMD and BMDL by model averaging is reviewed, as well as how data from more complex designs can be used to determine BMD and BMDL. Furthermore, one example touches upon time-to-

event data, a data type that has been overlooked in BMD context, while the last example illustrates how SSD analysis can be based on BMDL rather than ED50 (or EDx, NOAEL or LOAEL). This gives more robust hazard concentrations, since the uncertainty is taken into account to a greater extent than before.

Based on a large simulation study of the different methods to estimate BMDL, we found that inverse regression does not work optimal for BMDL estimation when looking at very small changes in the response. The delta method, however, provides reasonable coverage probabilities, especially for the "extra" definition where the upper limit is utilized in the definition. On the other hand, the results of the bootstrap turned out surprisingly well in almost every scenario. In general, the simulations showed that it is wise to have such high doses included in the study that the upper limit of the dose-response relationship can be determined. Especially if the upper limit is directly used in the definition of BMD.

#### Conclusion:

The BMD methodology has great potential in ecotoxicology, including as a method for assessing environmental effects in pesticide regulation. A freely available R-package that addresses several of the issues found in a review of the literature within the BMD methods has been developed.

This report documents the package that, among other things, facilitate BMDL estimation by the bootstrap method, which in a simulation study resulted in consistently good results (coverage probabilities very close to the nominal level). The project has also shown that the BMD methodology is flexible enough to handle analytical problems in ecotoxicology, also for other types of data than those where the method is already applied today. This is illustrated through common data examples from ecotoxicology. An innovation that can improve the interpretation of the classical SSD analyzes is that you can now use BMDL instead of ED50 or NOAEL or LOAEL.



## 3. Indledning

En risikovurdering af et kemikalie i forhold til mennesker og miljø består overordnet set af et mål for giftighed, en "farligheds-vurdering", sammenholdt med et mål for eksponering til kemikaliet. I det første trin identificerer man om et kemikalie kan udgøre en specifik form for giftighed (akut giftighed, kræftfremkaldende, hormonforstyrrende etc.). Dernæst identificerer man ved hvilke type eksponering og hvilke dosering giftige virkninger opstår, og endelig sammenholder man disse informationer med en vurdering af eksponeringen. Processen omkring risikohåndtering og regulering er en politisk proces hvori det defineres hvor stor en risiko der er acceptabel for det givne kemikalie og de mennesker/det miljø man vil beskytte.

Som udgangspunkt for en farlighedsvurdering, kan man finde en "dose of departure" (DOD), som den laveste dosis eller koncentration af et kemikalie, der afviger fra responsen af de ubehandlede. Denne DOD kan tjene som mulige udgangspunkt for den efterfølgende vurdering af risiko, når den sammenholdes med eksponerings scenarier. Afhængigt af eksponeringsscenarioet vil ekstrapoleringen bestå i, at DOD modificeres med en sikkerhedsfaktor (sammensat af delelementer der afhænger af karakteren af data) for at ende ud i en referenceværdi/grænseværdi for det givne kemikalie anvendt på en given måde (Hardy et al. 2017).

Nærværende dokument fokuserer på benchmark dose (BMD) metoder der bruges til på baggrund af en dosis-respons model at finde DOD til fareligheds-vurderingen i risikovurderingsprocessen. Med den seneste opdatering af guidelines fra EFSA vedrørende BMD, understreges anbefalingen af BMD frem for NOAEL hvor muligt (Hardy et al. 2017). BMD metoden bruges til, på basis af regressioner, at definere den laveste dosis eller koncentration der afviger med en vis størrelse fra responsen af de ubehandlede i et klassisk dosis-respons set up. Metoden anvender ikke parvise sammenligninger som når der findes NOAEL/ LOAEL.

EFSA anbefaler ikke alene at rapportere BMD men også konfidensintervallet omkring BMD da nedre grænse potentielt skal bruges som udgangspunkt for grænseværdisætning. Den foretrukne metode til beregning af BMD og det tilhørende konfidensinterval er nu blevet modelgennemsnit (Hardy et al. 2017; Slob 2018). Disse anbefalinger understøttes af anbefalinger fra US EPA (US Environmental Protection Agency 2012)

BMD-metoden blev første gang introduceret af Crump for binomiale respons data (Crump 1984) med en efterfølgende udvidelse til kontinuerte data (Crump 1995) og er en metode der bruger en dosis-responsmodel til at udlede modelbaserede DOD'er til lavdosis ekstrapolering (Edler et al. 2002; Wheeler et al. 2015; Wheeler and Bailer 2013). BMD er defineret som den dosis, der resulterer i en lille på forhånd fastlagt "acceptabel" stigning, benchmark response (BMR), i sandsynligheden for et uønsket event i forhold til baggrundssandsynligheden (figur 1). I praksis bruges ofte den nedre grænse for konfidensintervallet for den estimerede BMD (BMDL) som DOD til grænseværdifastsættelse.

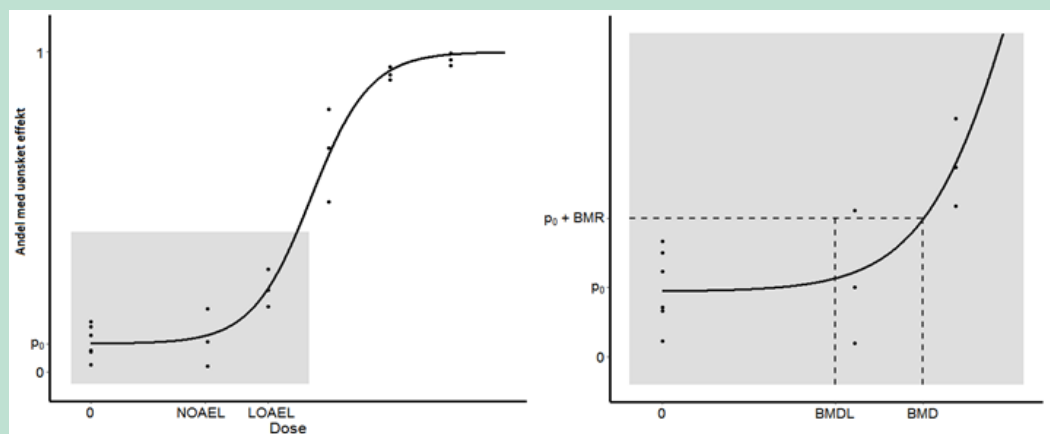
BMD er et alternativ til No Observed Adverse Effect Level (NOAEL) og Lowest Observed Adverse Effect Level (LOAEL), som har flere uønskede egenskaber (Hardy et al. 2017; Organisation for Economic Co-operation and Development 2012; Risk Assessment Forum and U.S. Environmental Protection Agency 2005). NOAEL er her defineret som den højest testede dosis/koncentration der ved parvise sammenligninger til kontrolgruppen ikke udviser nogen effekt. Der er mange fordele ved BMD metoden frem for brug af NOAEL:

- NOAEL metoden udnytter kun to doser ad gangen, mens BMD metoden udnytter hele dosis-respons kurven i analysen. Det betyder, at der for et givet antal organismer kan udtrækkes mere information i BMD-metoden end ved brug af NOAEL.

- BMD tillader (og bygger på muligheden for) at kvantificere præcisionen af estimatet, hvad der ikke er en mulighed for NOAEL/LOAEL.
- Gode eksperimentelle design med velplacerede doser og nok gentagelser belønnes med BMD-metoden. For et stigende antal observationer i datasættet vil BMD holde sig på et konstant niveau, mens usikkerheden omkring estimatet vil blive mindre og mindre og BMDL dermed vil rykke tættere og tættere på BMD. I modsætning hertil vil et øget størrelse datasæt medføre at NOAEL bliver mindre og mindre, da et større datasæt muliggør detektion af mindre forskelle. Dermed kan valg af design også have vidt forskellige konsekvenser alt efter om der bruges NOAEL eller BMD(L).
- BMD(L) kan ofte bestemmes selv når der ikke observeres et NOAEL, fordi BMDL gør brug af hele datasættet (Izadi, Grundy, and Bose 2012).
- BMD-estimering danner grundlag for en mere raffineret DOD-afledning i tilfælde af en stor dosisafstand mellem NOAEL og LOAEL da BMD ikke behøver at være en dosis i eksperimentet. BMD er derfor ikke lige så afhængig af placeringen af doser.
- Responsniveauet hørende til NOAEL er ikke det samme mellem studier. For BMD beholder responsniveauet den samme fortolkning for alle studier. Mens BMDL vil være dosen for hvilken vi med 95% sikkerhed kan sige at responsniveauet ikke overskrider det fastlagte.

Omvendt lider NOAEL ikke under nogen form for bias som konsekvens af en forkert antagelse omkring dosis-responsmodellen. Det kan være vanskeligt at estimere en dosis-respons-model til relativt enkle studier, hvor der kun indgår en enkelt eller to eksponerede grupper ud over kontrollen. Desuden kan doserne testes uafhængigt af tilstedeværelsen af de andre doser i NOAEL tilgangen, hvilket tillader en rekursiv testning hvor næste dosis kun testes hvis det ønskede resultat ikke er opnået. Derimod skal alle doser bruges til estimation af dosis-respons kurven der ligger til grund for BMD metoden.

Til trods for anbefalingen af BMD frem for NOAEL er udmeldingen imidlertid fra EFSA at det ikke nødvendig med en generel reevaluering af tidligere vurderinger der bygger på NOAEL (Hardy et al. 2017).



**FIGUR 1.** Benchmark dosen (BMD) defineret i termer af "additional risk". Til venstre ses hele dosis-respons kurven, mens den nedre del af kurven, illustreret ved den grå baggrund, er forstørret på figuren til højre. BMD metoden er en metode til farlighedsvurdering hvor den nedre ende af dosis-respons kurven benyttes. BMD er den dosis der resulterer i en på forhånd fastlagt lille ændring i respons (BMR) i forhold til baggrunds niveauet,  $p_0$ . Den nedre grænse af konfidensintervallet for BMD, kaldet benchmark dose nedre grænse (BMDL), er dosen der bruges som udgangspunkt for fastsættelse af grænseværdier. Modsat No Adverse Effect Level (NOAEL) og Lowest Adverse Effect Level (LOAEL), så behøver hverken BMD eller BMDL at være en dosis anvendt i forsøget.

BMD-metoden ligger sig tæt op af effektiv dosis (ED/EC/LD) bestemmelse, hvor der er tale om doser/ koncentrationer der resulterer i en på forhånd fastlagt ændring i responsen på relativ skala i forhold til både øvre og nedre grænse, hvor BMD som oftest ser på ændringer i forhold til kontrolgruppen alene (med enkelte undtagelser). Som eksempel kan det nævnes at for binomiale respons data (i tilfældet hvor alle individer påvirkes ved høje doser) er BMD bestemt ved "excess risk" definitionen og et BMR på 10% det samme som ED<sub>10</sub>. Største forskel ligger dog i at det med BMD metoden er BMDL og ikke BMD der bruges som DOD, hvorimod det for effektive doser er punkttestimatet (for eksempel ED<sub>10</sub>) er bruges.

Organisationen for økonomisk samarbejde og udvikling (OECD) samt flere regulerende myndigheder i forskellige lande samt i EU anerkender BMD-konceptet som værende mere videnskabelig underbygget end NOAEL/LOAEL metoderne og har endvidere lavet vejledninger til brugen af BMD (Davis, Gift, and Zhao 2011; Hardy et al. 2017; Organisation for Economic Cooperation and Development 2012). Da BMD metoden beskriver usikkerheden i risikovurderingen bedre end NOAEL (World Health Organization 2014), er det den anbefalede metode i risikovurderingen af kemikalier (European Chemicals Agency 2017; International programme on chemical safety 2009), plantebeskyttelsesmidler (Hardy et al. 2017; US Environmental Protection Agency 2012), urenheder i lægemidler (International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use 2014) og kosmetik (Scientific Committee on Consumer Safety 2016). Metoden har dog stadig ikke opnået en samlet accept som den foretrukne fremgangsmåde for risikovurderinger inden for økotoxikologi. Det kan muligvis skyldes at bestemmelsen af BMD kan forekomme mere kompliceret (og skræmmende) end den enkelhed der ligger bag NOAEL/LOAEL. En alternativ forklaring kan være at brugen af BMDL som DOD, dvs. indbygningen af usikkerheden i DOD, kan reducere DOD dramatisk i nogle regulatoriske studier på grund af begrænset gruppe- og observationsantal samt valg af doser, med begrænsede muligheder for optimering på grund af etiske betænkeligheder med hensyn til dyreforsøg. Til gengæld bruges de relaterede EC<sub>x</sub>/LC<sub>x</sub> som oftest inden for økotoxikologi.

På trods af anbefalinger fra myndigheder som OECD (Organisation for Economic Cooperation and Development 2014), EFSA (Bresson et al. 2009; Hardy et al. 2017) og US EPA (US Environmental Protection Agency 2012) om at bruge BMD / BMDL er metoden endnu ikke blevet en integreret del af risikovurderinger uden for Nordamerika og i langt mindre grad til risikovurdering inden for økotoxikologi, såsom risikovurdering af den økotoxikologiske virkning af bekæmpelsesmidler.

Anvendelse af BMD og fortolkningen af resultater er ikke trivielt og kræver statistiske programmeringsfærdigheder eller specialiseret software. Bl.a. efterspørger EFSA i deres seneste anbefalinger om BMD specialiseret software der kan beregne BMD baseret på model-gennemsnit (Hardy et al. 2017). Forskellige typer af data kræver forskellige definitioner af BMD og ligeledes kan usikkerheden i data indbygges i BMD på forskellig vis. Der findes kun få software programmer til formålet og ingen af de eksisterende løsninger er særlig fleksible hvad angår valg af dosis-respons-modeller og typer af data.

Formålet med dette projekt har således været todelt:

- At forbedre videns grundlaget der ligger til grund for den aktuelle pesticidregulering gennem bedre udnyttelse af dataressourcer, herunder identifikation af variationskilder, ved at introducere BMD metoden. Disse evidensbaserede forbedringer kan i sidste ende føre til en mere datadrevet og objektiv reduktion i miljørisiko og påvirkning af pesticider.
- At bygge bro mellem det anvendte aspekt og den teoretiske udvikling af BMD-metoder og forsyne økotoxikologer, forskere og myndigheder med tilgængelige og praktiske værktøjer gennem implementering af metoderne i tilgængelige open source-softwaremoduler.

Denne rapport indeholder en gennemgang af det eksisterende litteratur inden for BMD meto-  
dologien, efterfulgt af en beskrivelse af og kort introduktion til den implementerede R-pakke til  
BMD bestemmelse. Derefter gennemgås seks eksempler på brug af BMD-metoden inden for  
økotoksikologi for at illustrere hvordan forskellige statistiske og data-relaterede problemstillin-  
ger kan håndteres. Til sidst viser vi resultaterne af et større simulationsstudie der sammenlig-  
ner forskellige metoder til bestemmelse af BMDL inden vi diskuterer projektets resultater.

# 4. Eksisterende litteratur på metoder omkring BMD

Dette afsnit indeholder resultaterne af et gennemgående litteraturstudie af metoder til BMD bestemmelse med fokus på udviklingen inden for det seneste årti.

## 4.1 Udledning/bestemmelse af BMD

BMD metoden kan groft sagt opsummeres i fire trin:

- 1) Valg af definition af BMD på baggrund af et på forhånd fastlagt BMR.
- 2) Fastsættelse af BMR, der definerer den lille ændring i forhold til kontrolgruppen.
- 3) Valg af en dosis-respons-model.
- 4) Estimering af BMD og den tilhørende nedre grænse (fra konfidensintervallet) BMDL, som er den nedre grænse for plausible BMD-værdier, når der tages hensyn til det eksperimentelle design og usikkerheden i responsværdier. BMDL bruges som DOD til afledning af grænseværdier.

I praksis går de første to trin hånd i hånd; mens fastlæggelsen af BMR, eller mere præcist, det numeriske og filosofiske udtryk for BMR, afhænger af hvordan BMD defineres, vil en relevant toksisk tærskelværdi eller effektområde antageligt blive defineret først (hvorfor vi udtrykker det som et "på forhånd fastlagt" BMR).

### 4.1.1 Valg af definition af BMD for et på forhånd fastlagt BMR

BMD er en afledt parameter fra dosis-respons-modellen der afhænger af det på forhånd fastlagte BMR. Lad  $f(\text{Dose}, \beta)$  være funktionen der beskriver sammenhængen mellem middelværdien i respons og dosis ved hjælp af parametervektoren,  $\beta$  (se figur 1).

#### **Definitioner for binomiale respons data**

Der findes to definitioner af BMD for binomiale responsdata,

$$BMR = \frac{f(BMD, \beta) - p_0}{1 - p_0}, \quad (1)$$

og

$$BMR = f(BMD, \beta) - p_0 \quad (2)$$

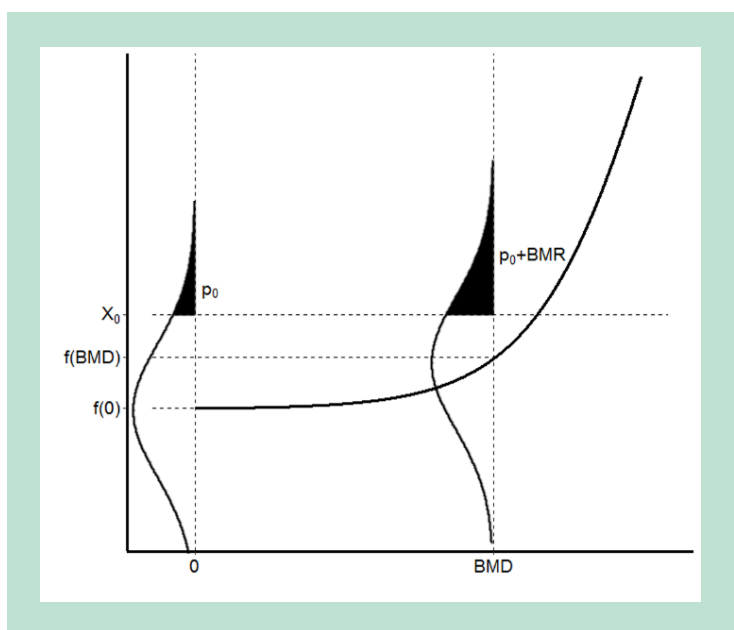
Definitionen i ligning 1 kaldes "excess risk" og definerer BMD, som den dosis der resulterer i en relativ ændring i sandsynligheden for et uønsket respons i forhold til baggrundsniveauet. Definitionen i ligning 2 kaldes "additional risk" og definerer BMD som dosis der resulterer i en absolut ændring i forhold til baggrundsniveauet (se figur 1). I ligningerne ovenfor er  $f$  dosis-respons-modellen og  $\beta$  er vektoren af parametre der indgår i dosis-respons-modellen.  $p_0$  repræsenterer baggrundssandsynligheden, dvs. andelen af kontrolgruppen der oplever en uønsket event. Baggrundsniveauet kan enten specificeres baseret på ekspertviden, estimeres fra kontrolgruppen alene eller fra dosis-respons-modellen. Både U.S. EPA og EFSA anbefaler brugen af "excess risk" definitionen.

#### **Definitioner for kontinuerte respons data**

Selvom det kan være fristende at oversætte (1) og (2) direkte til kontinuerte respons data, er det ikke så ligetil. Definitionen for kontinuerte respons data der bedst ligner de ovennævnte definitioner for binomiale respons data er den såkaldte hybrid metode (Budtz-Jørgensen,

Keiding, and Grandjean 2001; Gaylor and Slikker 1990, 2004; Kodell and West 1993). Her defineres BMD som den dosis der svarer til en på forhånd fastlagt stigning i risikoen / sandsynligheden for at falde under (eller overskride) en bestemt tærskelværdi på responskalaen, hvor der bruges et cut-off til at opdele respons værdierne i intervaller med acceptable og uønskede værdier. Andre definitioner er baseret på absolutte eller relative ændringer på den originale skala (Budtz-Jørgensen et al. 2013; Slob 2017; Slob and Pieters 1998). Alternative definitioner er blevet foreslået men har ikke vundet indpas (Sand et al. 2006; Slikker et al. 2004). For eksempel er det blevet foreslået at definere BMD som den dosis hvor hældningen af dosis-responsfunktionen ændrer sig mest (Sand et al. 2006).

Hybridtilgangen ligner definitionen fra binomiale respons data i brugen af en sandsynlighed for et uønsket event. I modsætning til binomiale data, hvor det uønskede event observeres direkte, kan baggrundsniveauet,  $p_0$ , for andelen af individer der oplever et uønsket event ikke estimeres direkte ud fra de kontinuerte data, men skal specificeres ved hjælp af et på forhånd specificeret cut-off (se figur 2).



**FIGUR 2.** Hybrid metoden til bestemmelse af benchmark doser for kontinuerte response data. Hybrid metoden definerer baggrundssandsynligheden for et uønsket event,  $p_0$ , som sandsynligheden for at baggrundspopulationene (kontrolgruppen) ligger over (under) et givet cut-off,  $x_0$ . Benchmark dosen (BMD) defineres som dosen der resulterer i et på forhånd fastlagt ændring, BMR, i sandsynligheden i forhold til baggrundssandsynligheden.

sandsynligheden i forhold til baggrundssandsynligheden.

For eksempel kan en percentil,  $x_0$ , i fordelingen for den ueksponerede population, dvs. kontrolgruppen, fungere som cut-off enten ved at bruge den empiriske fordeling eller ved at inkludere fordelingsantagelser. Antages det for eksempel at responsværdierne for kontrolgruppen stammer fra en normal fordeling, kan baggrundeniveauet/baggrundssandsynligheden specificeres som følger

$$p_0 = 1 - \Phi\left(\frac{x_0 - f(0, \beta)}{\sigma}\right), \quad (3)$$

for en dosis-respons-model beskrevet ved funktionen funktion  $f$  og parametervektoren  $\beta$ .  $\Phi$  angiver den kumulative fordelingsfunktion for en standard normalfordeling. Baggrundsniveauet kan så, ligesom for de binomiale respons data, defineres baseret på kontrol gruppen, dvs. den samtidige kontrol, historiske data eller alternativt en blanding af begge. I tilfælde af varianshomogetitet kan alle residualer fra dosis-respons-modellen alternativt anvendes til at estimere standard afvigelsen. Denne tilgang antager, at standard afvigelsen er uafhængig af dosis, men kan generaliseres til ethvert tilfælde hvor standard afvigelsen kan estimeres ved anvendelse af data fra alle doser, for eksempel eksponentielt voksende standard afvigelser som funktion af dosis.

For nogle kontinuerte respons variable kan et specifikt niveau betragtes som værende uønsket og kan dermed agere som et naturligt cut-off som alternativ til brug af percentiler.

På samme måde som for binomiale respons data definerer hybridmetoden BMD i form af "excess risk", altså som dosen der resulterer i en på forhånd fastlagt lille relativ ændring fra baggrundsniveauet:

$$BMR = \frac{1 - \Phi\left(\frac{x_0 - f(BMD, \beta)}{\sigma}\right) - p_0}{1 - p_0} \quad (4)$$

Ligeledes kan hybridmetoden definere BMD som dosen resulterende i en absolut ændring ("additional risk"):

$$BMR = 1 - \Phi\left(\frac{x_0 - f(BMD, \beta)}{\sigma}\right) - p_0 \quad (5)$$

Hybridmetoden er ret følsom over for fordelingen af responsen, især fordi halerne af fordelingen kan variere meget. Derfor kan den underliggende fordelingsantagelse for data rapporteret som summarydata ved hjælp af middelværdi og standard afvigelse have stor indflydelse på den estimerede BMD og særligt BMDL der jo afhænger stærkt af hvad der foregår i halerne af fordelingen (Shao, Gift, and Setzer 2013). Robust estimation, dvs. metoder, der er mindre afhængige af varianshomogenitet og ekstreme værdier, af dosis-responsmodellen kan måske hjælpe med at overvinde nogle af de udfordringer, der er forbundet med varians inhomogenitet, som måske ellers ville blive håndteres gennem en antagelse om log-normal-fordelt data.

Som et alternativ til hybrid metoden kan BMR definere en relativ ændring (sædvanligvis en procentvis ændring) i den gennemsnitlige respons i forhold til respons niveauet for kontrollen. Her ignoreres variationen i de målte responsværdier helt i definitionen. Denne tilgang kaldes sommetider bestemmelse af den kritiske effektstørrelse (Slob 2017). Wheeler og Bailer (Wheeler and Bailer 2009b) benyttede tælle data på reproduktion og definerede BMD som den dosis der reducerer størrelsen med en bestemt andel af baggrundsniveauet. En lignende definition er blevet foreslået inden for bestemmelse af ED-værdier for familien af generaliserede lineære modeller (Bailer and Oris 1997).

Et uafklaret problem i forbindelse med de relative definitioner er, at der ikke er enighed om relevante ændringer i responsen, dvs. den 'toksiske effekts størrelse' af de forskellige individuelle biologiske responser. For eksempel kan et fald i kropsvægt på 20% betragtes som toksikologisk relevant, mens en ændring på 20% i hormonkoncentration kan være biologisk plausibel i en behandlingsgruppe. Derfor er der behov for at diskutere valg af relevante størrelser af BMR mellem interessenter i de relevante reguleringssektorer. Bestemmelse af "den kritiske effektstørrelse" har derudover sine begrænsninger da variationen i kontrolgrupper kan variere med flere størrelsesordner fra et forsøg til et andet (Zeller, Duran-Pacheco, and Guérard 2017). Det er derfor vigtigt med en fælles forståelse og konsensus om normal variation på tværs af forskellige respons variable.

Selvom ekspertvurderinger kan anvendes til at kvantificere den kritiske effektstørrelse for forskellige respons variable kan det være svært at kvantificere dem på en måde hvor de kan betragtes som havende den samme betydning mellem respons variable. Et forsøg på at imødekomme dette omfatter et forslag om at skalere den kritiske effektstørrelse baseret på den maksimale respons og variationen inden for gruppen (Slob 2017).

Endelig foreslog Wheeler et al. (Wheeler et al. 2017), at definere BMD ved at anvende percentil-regression. Ved at modellere dosis-percentil forholdet definerer man sandsynligheden for en uønsket effekt implicit og imødekommer endvidere situationen, hvor både middelværdi og varians afhænger af dosis.

Respons variable af andre typer end binomial eller kontinuerte data er ikke almindelige i BMD-litteraturen. Chen og Chen (Chen and Chen 2014) beregnede BMD for ordnede kategoriske data (f.eks. ingen, lav, middel, høj effekt) ved at definere BMD som et vægtet gennemsnit af de estimerede BMD-værdier ved hjælp af cut-off på forskellige intensitetsniveauer. Wheeler og Bailer (Wheeler and Bailer 2009b) betragtede BMD for tælledata. Parallele metodeudviklinger for ED bestemmelse baseret på tælledata kan også være nyttige til BMD estimering (Delignette-Muller et al. 2014; Zhang, Bailer, and Oris 2012b, 2012a). Så vidt vi ved, har der ikke været nogen generel definition af BMD for tid-til-event data for eksempel tid til død eller tid til første tumor, der begge er vigtige parametre i toksikologi. Morales og Ryan (Morales and Ryan 2005) beregnede dog BMD for tid-til-event data i et specialtilfælde indenfor epidemiologiske data ved hjælp af poisson regression.

#### 4.1.2 Fastsættelse af BMR

For en relevant risikovurdering er det nødvendigt at definere et passende BMR. Ideelt set er det et BMR baseret på ekspertens biologiske og toksikologiske viden om testsystemet og/eller eksperimentet sammenholdt med det beskyttelses niveau man ønsker inden for en given risikovurdering. Dog bør beslutningen bygge på en kombination af etablerede konventioner og anerkendte statistiske metoder.

Almindelige værdier for BMR er 5% og 10% ændringer i responser for henholdsvis kontinuerte og binomial data (European Food Safety Authority scientific committee 2016; US Environmental Protection Agency 2012). Dette giver rigtig god mening for binomiale dosis-respons data, hvor responset angives som proportioner mellem 0 og 1 (0% og 100%), men mindre mening for kontinuerte data, hvor der ikke er en universel skala for alle typer af respons (Izadi, Grundy, and Bose 2012).

U.S. EPA anbefaler at der vælges et BMR baseret for det specifikke datasæt, i den nedre ende af det observerede dosisområde. For binomiale respons data er en BMR på 10% "excess risk" normalt inden for området og er et almindeligt valg for at sammenligne styrker (virkning) (baseret på BMD) på tværs af kemikalier og respons variable, hvorfor US EPA anbefaler som minimum altid at bruge et BMR på 10% med "excess risk" definitionen. For epidemiologiske datasæt kan et BMR på 10% dog ofte indebære en ekstrapolering opad, da der oftest ikke er doser i data der er associeret med så store ændringer fra kontrollen, hvorfor et lavere niveau, for eksempel 1%, kan være mere passende (US Environmental Protection Agency 2012).

For kontinuerte data anbefaler US EPA rapportering af resultater efter brug af både hybrid metoden og bestemmelse af den kritiske effektstørrelse. For førstnævnte anbefales det at bestemme BMD svarende til en ændring i den gennemsnitlige respons på 1 standard afvigelse for kontrolgruppen. Dermed kan en estimeret BMD baseres på hybridmetoden med en BMR svarende til 5% eller 10%. Andre almindelige cut-off værdier inkluderer en baggrundssandsynlighed ( $p_0$ ) på 1% sammen med et BMR på 5 eller 10% (Yang, Allen, and Thomas 2007).

EFSA anbefaler at bestemme den kritiske effektstørrelse svarende til BMR på 5% (Hardy et al. 2017).

Valget af BMR påvirker ikke kun BMD og BMDL estimererne. De sikkerhedsfaktorer der anvendes til at ekstrapolere fra BMDL til en grænseværdi bør også justeres når data muliggør anvendelse af et mindre BMR (Haber et al. 2018). Der mangler dog stadig forskning omkring hvilke implikationer størrelsen af BMR har på ekstrapoleringen til grænseværdibestemmelse.

#### 4.1.3 Valg af dosis-respons-model

Da de fleste biologiske respons er positive bør de statistiske modeller bruges til at bestemme BMD være i overensstemmelse med denne begrænsning (US Environmental Protection Agency 2012). Derudover skal den estimerede dosis-respons trend være monoton, dvs. kun vokse (eller aftage) i hele dosisområdet. Der kan sagtens afviges fra begge disse krav, men BMD-estimering ligger så uden for det typiske anvendelsesområde og kan involvere ekstra begrænsninger og udfordringer med hensyn til fortolkninger. For eksempel vil bestemmelse af



BMD for et stof der inducerer et ikke-monotont dosis-responsforløb i form af en bifasisk dosis-respons-kurve, betegnet hormesis, kræve en beslutning om hvad der skal bruges som baggrunds niveau; det estimerede niveau for kontrolgruppen eller det maksimale niveau (Bailer and Oris 1997, 2000). Det kan i det tilfælde også være nødvendigt at begrænse fokus til kun en del af kurven, da mere end en dosis kan resultere i den på forhånd fastlagte ændring BMR fra baggrunds niveauet.

### **Parametriske modeller**

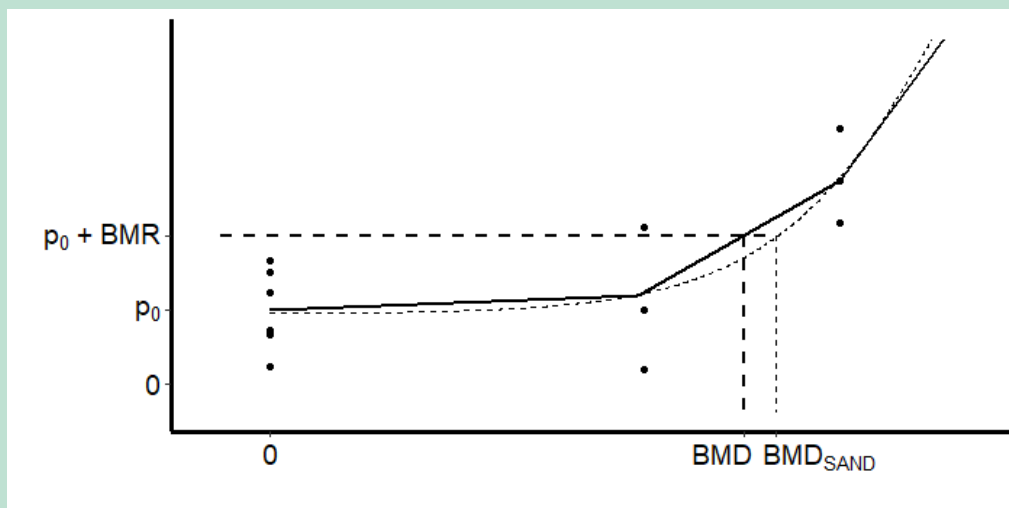
Modeller der indeholder et begrænset antal parametre, så som multistage, log-logistik, Weibull og gamma modeller, er ofte glimrende til at beskrive de fleste dosis-respons forløb (Buckley and Piegorsch 2008; Buckley, Piegorsch, and West 2009; Fox, Hogan, and Davis 2017; Kerns 2017; Kopylev and Fox 2009; Moon et al. 2013; Peng, Robichaud, and Alsubie 2015; Shao and Small 2011; Taft and Hines 2012; West et al. 2012; Wheeler and Bailer 2009a). Disse modeller er biologisk plausible og er til dels valideret, og de har vist sig at forudsige responsforløb i mange tilfælde. Generelt er parametriske modeller lette at implementere, de kræver mindre tid til beregning end deres semi- og ikke-parametriske modstykker og giver parametre, som er let at fortolke.

Namata et al. (Namata et al. 2008) udvidede de klassiske parametriske modeller med en ny og fleksibel familie af dosis-respons-modeller, de såkaldte fractional polynomiske log-logistiske modeller, som gør det muligt at modellere en bredere vifte af dosis-responsmønstre end de mere klassiske modeller. For nylig foreslog Das (Das 2018) at bruge en familie af parametriske dosis-respons-modeller for at spænde over et endnu bredere interval af dosis-responsmønstre, herunder lineære og supra-lineære forhold, som kan være svære at fange med de klassiske parametriske dosis-respons-modeller. I praksis er anvendelsen af familien af modeller imidlertid begrænset af det høje antal parametre.

### **Semi- og ikke-parametriske modeller**

Som beskrevet ovenfor kan de fleste, men ikke alle, dosis-respons-mønstre beskrives godt af de klassiske dosis-respons-modeller. For at imødekomme alle mønstre, har flere forfattere foreslået at man anvender semi- eller ikke-parametriske modellering (Guha et al. 2013; Lin, Piegorsch, and Bhattacharya 2015; Piegorsch et al. 2012, 2014), hvilket giver større fleksibilitet i beskrivelsen af dosis-responsmønstre. Dette kan dog også indebære en højere risiko for overfitting sammenlignet med de parametriske modeller. Das (Das 2018) viste lignende fit af hans familie af parametriske modeller sammenlignet med semi-parametriske modellering. En mere grundig sammenligning af bias af BMD og BMDL fundet på baggrund af de semi-parametriske, parametriske og de fractional polynomiske log-logistiske modeller mangler dog stadig.

Piegorsch et al. (Piegorsch et al. 2012, 2014) og Lin et al. (Lin, Piegorsch, and Bhattacharya 2015) foreslog en lineær interpolationsmetode der sikrer monotone modeller. Deres tilgang havde dog problemer med bias og dækningsgraden på konfidensintervallerne (og dermed BMDL) ved små sample data størrelser og konkave modeller på grund af brugen af lineær interpolation til at fange noget der ikke er lineært. Endelig synes metoden kun at være fordelagtig når man øger antallet af observationer, specielt hvis de blev fordelt over flere doser, da dette muliggjorde at den stykkevis lineære funktion i højere grad kunne fange en krumning. Det skal bemærkes, at flere af de foreslåede semi- og ikke-parametriske modelleringsstrategier er konservative i deres bestemmelse af BMDL, dvs. de resulterer i en noget lavere BMDL end nødvendigt for at opnå den ønskede dækningsgrad (Piegorsch et al. 2014; Wheeler and Bailer 2012), se figur 3. Det kan begrænse deres generelle anvendelse. For dosis-responsforløb der ikke er godt beskrevet med de tilgængelige parametriske modeller, kan de dog udgøre et alternativ.



**FIGUR 3.** Semi-parametrisk dosis-respons-model baseret på lineær interpolation mellem doser. En lineær interpolation metode kræver et højt antal doser for at beskrive et dosis-respons mønster med stor krumning. Ellers vil den estimerede benchmark dose (BMD) være for lav for konvekse sammenhænge og for høj for konkave dosis-respons sammenhænge og dermed resultere i bias.

#### **Bayesianske metoder**

En anden udvikling inden for dosis-respons-modellering er at anvende bayesianske modeller, hvor fordelingsmæssige antagelser om model-parametrene muliggør en beregningsmæssigt meget fleksibel og generel modelleringsmetode. Wheeler og Bailer (Wheeler and Bailer 2012) foreslog en bayesiansk generalisering af probit-regressionsmodellen. Simuleringer viste at passende placeringen af "knodepunkterne" i denne semi-parametriske metode var væsentlig for at opnå en nominal dækning, hvorfor validering af knode placeringen muligvis er et krav for brugen af denne metode. Guha et al. (Guha et al. 2013) foreslog to bayesianske ikke-parametriske metoder, der førte til ikke-negative, glatte monotone dosis-responskurver. Bayesianske metoder kan måske være nyttige hvis historiske data skal inkorporeres i analysen. Imidlertid er dosis-responsanalyse måske ikke det mest lovende felt for bayesianske metoder, da frekventistiske parametriske metoder er vel-tilpasset og veletableret for denne type data.

#### **4.1.4 Estimering af BMDL**

BMDL afhænger af den valgte dosis-responsmodel og af den fordeling der antages for respons data. Flere fremgangsmåder er blevet foreslået til at estimere BMDL. Generelt kan de inddeles i tre typer: 1) Asymptotiske metoder, der afhænger af antagelser om hvordan fordelingen af parametrene i modellen opfører sig for store datasæt for at udlede konfidensintervallet på en afledt parameter (dvs. BMD). 2) Invers regression, der estimerer usikkerhed på respons-skalaen. 3) Bootstrapping, der er baseret på resampling, dvs. simulation af lignende data ud fra observerede data. De fleste af de foreslåede metoder er blevet evalueret ved hjælp af simulationer, men en omfattende sammenligning af alle metoder mangler stadig for at der kunne udpeges en optimal tilgang.

#### **Asymptotiske metoder**

En simpel og ofte anvendt tilgang er baseret på asymptotiske Wald konfidensintervaller som let kan kombineres med en transformation for enten at opnå normalitet eller for at sikre meningsfulde (dvs. positive) estimater af BMDL (Buckley, Piegorsch, and West 2009; Moon et al. 2013; Namata et al. 2008; West et al. 2012). Wald konfidensintervaller for afledte parametre,

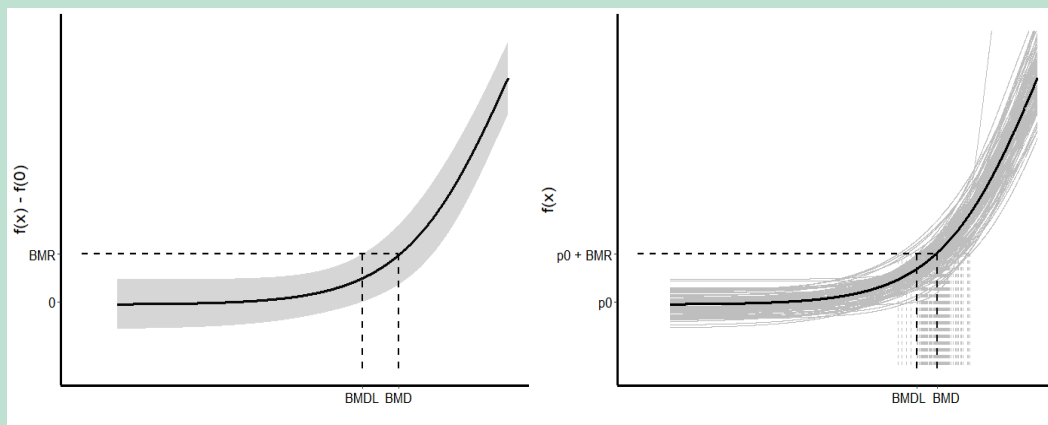
som BMD, findes ved at kombinere information om parametres fordelinger ved hjælp af delta-metoden. Alternativt kan modeller reparametriseres så en af modelparametrene bliver BMD. Derved findes konfidensintervallet indirekte som led i model estimationen.

En anden tilgang, der også er baseret på asymptotiske resultater, er at bruge profil-likelihood intervaller (Fox, Hogan, and Davis 2017; Izadi, Grundy, and Bose 2012; Ringblom, Johanson, and Öberg 2014; Yu and Catalano 2005). Profil-likelihood-metoden reducerer likelihood-funktionen til en funktion af en enkelt parameter af interesse, for eksempel BMD. Profil-likelihood-funktionen maksimeres så for at finde parameterestimatet, men kan også bruges til at estimere grænserne for det interval, for hvilket sandsynligheden rammer percentilerne af interesse for den tilhørende  $\chi^2$ -fordeling, som definerer profil-konfidensintervallet. Denne tilgang involverer normalt en omparametrisering af modellen for at inkludere BMD som en modelparameter. Det er blevet vist at profil-konfidensintervallet giver bedre dækning end de delta-metode-baserede Wald-konfidensintervaller, dvs. at det oftere indeholder den ukendte parameter med 95% sandsynlighed (Moerbeek, Piersma, and Slob 2004; Slob 2014b; US Environmental Protection Agency 2012). Man har imidlertid også vist, at dækningen af ensidige profil-likelihood intervaller for et BMD-estimat kan føre til en fejlagtig BMDL, når den sande værdi af en eller flere parametre i dosis-respons-modellen er på eller under grænsen for parameterområdet (Kopylev and Fox 2009).

### ***Invers regression***

Invers regression kan bruges på et konfidensbånd for kurven beskrivende ændringen i respons i forhold til baggrundsniveauet. Idéen er, at da responset er kendt, kan den tilhørende dosis findes (Demidenko et al. 2013). I praksis gøres dette ved at løse den estimerede respons-ændrings-funktion for dosis. Ved at bruge omvendt regression på konfidensbåndet bestemmes BMDL (Buckley, Piegorsch, and West 2009; Fang, Piegorsch, and Barnes 2015) (figur 4, venstre).

En udvidet version af dette ville være at bruge percentil-regression (Wheeler, Shao, and Bailer 2015), hvor percentiler i sig selv er modelleres som funktion af dosis. Wheeler et al. (Wheeler et al. 2017) foreslog en bayesiansk percentil-regressions metode hvor fleksible polynomier bruges til at modellere dosis-percentil kurven. Percentil-regression gør det muligt at estimere BMDL uden at antage en specifik fordeling af respons variabelen. Derudover lægges der mindre vægt på fordelingsantagelsen hvis percentil-regression også anvendes til at modellere medianen. Det kræver imidlertid store datasæt med mange gentagelser at få en robust estimering af percentilerne (Wheeler, Shao, and Bailer 2015) hvilket kan begrænse anvendeligheden af denne metode.



**FIGUR 4.** Estimation af benchmark dose nedre grænse ved brug af invers regression eller bootstrap. Venstre: Med invers regression defineres benchmark dose nedre grænse (BMDL) som den dosis der resulterer i den øvre grænse af konfidensbåndet for ændringen i respons i forhold til baggrundsniveauet rammer den på forhånd fastlagte ændring BMR. Højre: Bootstrapping. For hvert bootstrap data sæt estimeres en dosis-respons-model og den tilhørende BMD-værdi findes. BMDL er defineret som 5%-percentilen i bootstrap-fordelingen for BMD.

### Bootstrap

Endelig er forskellige bootstrap-strategier blevet foreslået i forbindelse med BMDL-estimering; både parametriske (Bailer and Oris 1997; Buckley, Piegorsch, and West 2009; Moon et al. 2013) og ikke-parametriske (Piegorsch et al. 2012, 2014). Kort fortalt er bootstrap en metode, hvor nye datasæt findes ved permutation fra det originale datasæt, eller ved simulation ud fra den originale model (parametriske), for at opnå fordelingen af den estimerede BMD (Figur 4, højre). BMDL kan igen findes ved en Wald-interval tilgang, men ved at erstatte den asymptotiske baserede standardfejl med en bootstrap-baseret standardfejl (Wu et al. 2006). Alternativt kan man finde BMDL som en percentil fra bootstrap-fordelingen (typisk 5% percentilen) eller bruge såkaldt bias-korrigeret bootstrap til at finde percentilen (Wheeler et al. 2015). Zhu og kolleger (Zhu et al. 2005; Zhu, Wang, and Jelsovsky 2007) foreslog en alternativ parametriske bootstrap-procedure, der kombinerede permutation af residualer med en et-trins estimering af parametrene for at slippe udenom de konvergensproblemer der ofte følger med gentagne estimationer af ikke-lineære dosis-respons-modeller som alle andre parametriske bootstrap-metoder er afhængige af.

Bootstrap-metoder kan anvendes i de fleste tilfælde, men kan være beregningstunge, da processen for hvert bootstrap-datasæt (typisk 1000 eller mere) kan være beregningsmæssigt kompleks (Buckley, Piegorsch, and West 2009; Kerns 2018; Moerbeek, Piersma, and Slob 2004).

## 4.2 Andre betragtninger

Mens det foregående afsnit gennemgik de grundlæggende trin til at udlede BMD, vil dette afsnit fokusere på udviklingen af metoder der er mere generelt anvendeligt for dosis-respons-modeller og anvendelser af BMD i praksis.

### 4.2.1 Design af eksperimenter

Dosis-respons-modeller drager generelt fordel af et øget antal doser og i mindre omfang et øget antal observationer pr. dosis. Der er dog, ud over statistiske overvejelser, også etiske betæneligheder og regulatoriske restriktioner der kan påvirke antallet af organismer der bruges i et eksperiment (Slob 2014a, 2014b). Det er derfor i flere sammenhænge blevet undersøgt hvordan designet af et eksperiment påvirker BMD-bestemmelsen.

Da binomiale respons data indeholder mindre information end kontinuerte respons data, er de førstnævnte mere kritiske når der skal vurderes hvor mange forsøgsorganismer der er behov for, for en robust estimering af BMD (Slob 2014b). For binomiale data resulterer forsøgsdesign der inkluderer en dosis med relativt høj forekomst af uønskede effekter, i en højere præcision af den estimerede BMD. Baseret på simuleringer og en efterfølgende Monte Carlo-analyse undersøgte Shao og Small (Shao and Small 2011) hvor et yderligere dosis niveau med fordel kunne inddrages for at minimere usikkerheden i BMD-estimerer. De fandt at inklusion af et relativt højt dosisniveau kan resultere i en højere præcision i BMD-estimering og en tilsvarende højere BMDL. I et efterfølgende omfattende simulationsstudie (Shao and Small 2012) fandt de imidlertid, at selv om en relativt høj dosis er nødvendig, resulterede en for høj maksimal dosis i en reduktion af præcision af den estimerede BMD, ligesom en for lav maksimal dosis konsekvent resulterede i en lavere præcision. Resten af doserne skulle spredes over den resulterende dosis-skala. Mere generelt har det gentagne gange vist sig at være fordelagtigt at have forsøg med en eller flere doser nær niveauet for den sande BMD (Piegorsch et al. 2014; Ringblom et al. 2017; US Environmental Protection Agency 2012). Det er klart, at et sådan forsøgsdesign kræver at tidligere resultater bruges til at optimere fordelingen af doser.

Der er en tendens til en højere præcision af estimeret BMD når et fast samlet antal forsøgsdyr fordeles over et højere antal doser (Shao and Small 2012; Slob et al. 2005). En anden god grund til at øge antallet af doser er at minimere risikoen for at gentage en undersøgelse på grund af tabet af en enkelt dosis gruppe eller et uheldigt valg af doser. Da mange toksikologiske respons testes samtidigt, kan deres individuelle dosis-respons ligeledes variere, hvorfor et optimalt fit vil afhænge af forskellige dosis-placeringer (Muri, Schlatter, and Brüscheiler 2009; Slob 2014b).

Muri et al. (Muri, Schlatter, and Brüscheiler 2009) gen-analyserede et stort antal forsøg for at kontrollere anvendeligheden af BMD-metoden på eksisterende data. De viste, at de fleste datasæt var af tilstrækkelig kvalitet til BMD-estimering, baseret på kriterierne for goodness-of-fit af dosis-respons-modellen, størrelsen af det estimerede BMD / BMDL-forhold, og om den estimerede BMD var inden for studiets dosis-område. Carr et al. (Carr et al. 2018) viste, at et højere antal observation/doser er nødvendigt for at estimere koncentrationen for median dødelighed ( $LC_{50}$ ) med en given præcision, hvis hældningen af dosis-respons-kurven er lav, end når den er stejl, samt når den sande  $LC_{50}$  ligger i yderkanten af koncentrationsområdet sammenlignet med når koncentrationerne er centreret omkring  $LC_{50}$ . En tilsvarende undersøgelse bør laves med fokus på BMD værdier i den lave ende af dosisområdet.

Kontinuerte respons data kan kræve andre design end binomiale respons data da kontinuerte respons data ikke har nogen "naturlig" øvre (eller nedre) grænse. Derfor kan det for denne type data være vigtigt at inkludere høje doser for at estimere den øvre (eller nedre) grænse (Ringblom, Johanson, and Öberg 2014).

I nogle typer toksicitets-studier, er risikokarakterisering ofte baseret på flere udfald, dvs. både binomiale og kontinuerte respons data. På den ene side resulterer dette i praktiske begrænsninger for hvordan valg af doser kan øge præcisionen af BMD, mens antallet af individer oprettholdes af dyrevelfærdsgrunde. Endvidere, på grund af praktiske begrænsninger, sker det ofte at nogle endpoints kun evalueres for en delmængde af dyrene. En sådan reduktion i data fører naturligvis til en vis reduktion i præcisionen. Imidlertid kan en velgennemtænkt strategi for valg af delmængde af data minimere tabet af præcision. Najita og Catalano (Najita and Catalano 2013) vurderede forskellige designs til "subsampling" i tilfældet med en fælles model for et kontinuert og et binært endpoint. De fandt, at for positivt korrelerede respons variable, vil oversampling af dyr med events for den binære variable til undersøgelse af det kontinuerte endpoint resultere i en præcision der er ligeså god som eller bedre end at vælge en simpel tilfældig delmængde. Det skyldes at en oversampling af dyr med events medfører en højere sandsynlighed for at observere data i yderkanterne af fordelingen for den kontinuerte respons variable, hvorfor fordelingen vil blive bedre beskrevet.

Mens metoder der bygger på parvise sammenligninger, såsom NOAEL og LOEAL, drager fordel af design med samme gruppestørrelser, er det ikke nødvendigvis tilfældet for BMD-metoden (Öberg 2010). I kontekst med dyreforsøg viste to undersøgelser (Ringblom et al. 2017), at det at ændre antallet af anvendte dyr mellem doser, mens det samme antal doser bibeholdes, kan reducere dyrets stress uden at ændre præcisionen for den estimerede BMD. En lignende konstatering blev rapporteret af Shao og Small (Shao and Small 2012), der kun fandt en begrænset effekt på BMD-præcision ved ubalanceret design, men en tendens til øget præcision ved at placere flere dyr i grupper med lavere dosering og færre i grupper med højere dosering, hvilket resulterede i højere dyrevelfærd (Öberg 2010).

#### 4.2.2 Model valg

Valg af model bør så vidt muligt være motiveret af de underliggende biologiske processer, men niveauet af biologiske detaljer, der kan inkorporeres i sådanne modeller er minimalt (Crump et al. 2010; World Health Organization 2014). For eksempel kan der for de fleste biologiske processer ses bort fra modeller der tillader en uendelig hældning. Ved (øko-) toksicitetsanalyser der bruger høje doser / koncentrationer, er den undersøgte toksiske respons meget sandsynligt resultatet af en blanding af forskellige biologiske processer som i forskellige grad kan bidrage til den samlede respons afhængigt af dosis. Derfor vil valg af en enkelt model der inkorporerer underliggende biologiske processer kræve en god biologisk forståelse af respons og virkemåde. I sidste ende fastlægges troværdigheden af en model til at interpolere og ekstrapolere hovedsageligt ved at se på hvordan den passer til de data der er til rådighed, og følgelig bliver modeludvælgelse i praksis en modeltilpasning og sammenligningsøvelse. Valget af dosis-respons-model kan være afgørende for de konklusioner der drages (West et al. 2012). For eksempel viste Ringblom et al. (Ringblom, Johanson, and Öberg 2014), at test for lack-of-fit og model-reduktion kan resultere i valg af modeller der giver for høje BMDL-værdier.

#### 4.2.3 Model-gennemsnit

På grund af de problemer der er forbundet med modeludvælgelse og det faktum at modeludvælgelse kan resultere i bias og konfidensintervaller der indeholder den ukendte parameter (BMD) mere eller mindre end det nominelle niveau (typisk 95%), foretrækkes model-gennemsnit af flere forfattere til at estimere både BMD og den tilhørende BMDL (Bailer, Noble, and Wheeler 2005; Faes et al. 2007; Jensen and Ritz 2015; Kang, Kodell, and Chen 2000; Moon et al. 2005; Namata et al. 2008; Wheeler and Bailer 2007, 2008). Desuden har det vist sig at ekstrapolering i lav-dosis området baseret på model-gennemsnit resulterer i mindre bias end lineær ekstrapolering ved at bruge DOD, fordi brug af dosis-respons-modeller giver mulighed for at fange en potentiel krumning (Wheeler and Bailer 2013; Wheeler, Shao, and Bailer 2015).

I model-gennemsnit kombineres resultaterne fra flere modeller vægtet efter hvor godt de beskriver data, dvs. modeller der beskriver data godt vægtes højere i estimeringen. Det gennemsnitlige BMD-estimat er typisk et vægtet gennemsnit af individuelle BMD-estimer fra K-kandidatmodeller:

$$\widehat{BMD}_{MA} = \sum_{k=1}^K w_k \widehat{BMD}_k$$

Alternativt foreslog Wheeler og Bailer (Wheeler and Bailer 2007) at tage gennemsnit af hele dosis-responskurver. Der mangler dog stadig en grundig sammenligning af de forskellige tilgange.

Modelvægte er ofte baseret på informationskriterier, for eksempel AIC eller BIC. Et almindeligt valg er at bruge vægte af formen:

$$w_k = \frac{\exp(-\Delta_k/2)}{\sum_{i=1}^K \exp(-\Delta_i/2)}$$

hvor  $\Delta_k$  er lig  $AIC_k$  eller  $AIC_k - \min_k AIC_k$  eller lignende med et andet informationskriterium (Buckland, Burnham, and Augustin 1997; Jensen and Ritz 2015; Moon et al. 2005, 2013; Peña et al. 2017; Piegorsch et al. 2013). Bayesiansk model-gennemsnit udnytter en anden form for

vægtning af modellerne (Fang et al. 2015; Shao and Small 2012) og det er muligt at inkorporere forudgående viden om mere biologiske plausible dosis-respons-mønstre i modelvægtene.

Wheeler og Bailer (Wheeler and Bailer 2009a) vurderede forskellige modeludvælgelsesstrategier i sammenligning med model-gennemsnit til BMD-estimering. De sammenlignede model-gennemsnit med en "bedste" model, udvalgt ved Pearsons  $\chi^2$  goodness-of-fit-test, og med en strategi der udvælger det mindste BMD-estimat og dets nedre grænse. I de fleste tilfælde fandt de, at der var mindre bias ved BMD fundet ved model-gennemsnit. Det resulterede i en BMDL forbundet med et konfidensinterval tættere på det nominelle niveau. Derudover fandt de, at det lille bias der blev observeret for de model-gennemsnits BMD-estimer svarede til biasen for BMD-estimer fra den model der genererede de simulerede data. For nylig fandt en empirisk sammenligning endvidere ens estimer af lavdosis ekstrapolering for en semi-parametrisk og en model-gennemsnits tilgang og at begge resulterede i konfidensintervaller tættere på det nominelle niveau end en enkelt parametrisk "bedste" model (Wheeler and Bailer 2013).

West et al. (West et al. 2012) undersøgte indflydelsen af model-usikkerhed på BMDL-estimering og viste, at BMDL ofte er associeret med et respons der overgår det BMR der sigtes mod, og altså er ikke-beskyttende, når estimer er baseret på en udvalgt "bedste model", der ignorerer usikkerhed forbundet med modeludvælgelse. Peña et al. (Peña et al. 2017) leverede simulationsbaserede sammenligninger af forskellige BMD bestemmelses og model-udvælgelsesstrategier, både en ikke-parametrisk tilgang og model-gennemsnit baseret på AIC og BIC vægte. De viste, at de to BMD-estimer baseret på model-gennemsnit fungerer bedre end de forskellige enkeltmodelstrategier med hensyn til mindre bias i BMD og BMR med en anelse bedre resultater ved brug af AIC-vægte. Wheeler og Bailer (Wheeler and Bailer 2012) leverede ligeledes simulationer der muliggjorde sammenligning af deres egen semi-parametriske tilgang med model-gennemsnit for et par scenarier og viste stort set ens resultater for de to metoder.

Mens model-gennemsnit generelt klarer sig bedre end at vælge en "bedste" model, kan model-gennemsnit omvendt resultere i alvorlig bias som gentagne gange vist i simulationer (Simmons et al. 2015; Wheeler and Bailer 2007, 2009a), hvis udvælgelse af kandidat modeller ikke overvejes nøje. F.eks. kan model-gennemsnit mislykkes, hvis den sande model er uden for eller på kanten af rummet af kandidat modeller (Das 2018; Wheeler and Bailer 2009a, 2012). Flere ting er blevet foreslået for at komme uden om denne problematik: Faes et al. (Faes et al. 2007) introducerede fraktionelle polynomier i et sæt af kandidatmodeller til model-gennemsnit. Noble et al. (Noble, Bailer, and Park 2009) foreslog at reducere sættet af kandidatmodeller til model-gennemsnit ved at ekskludere modeller med en lav a posteriori sandsynlighed sammenlignet med den bedste model fra kandidatsættet. En yderligere reduktion blev foretaget ved at anvende "Occams razor" der kort fortalt går ud på at ekskludere modeller der er lige så dårligere end eller lige så gode som modeller de er indeholdt i. En anden fremgangsmåde til reduktion af kandidatmodeller er at bruge et indeks der afbalancerer godness-of-fit med diversitet af modeller (Kim, Kodell, and Moon 2014). I øjeblikket mangler der omfattende sammenligninger mellem ovenstående fremgangsmåder, og det er stadig vanskeligt at beslutte, hvilken fremgangsmåde der kan anbefales under hvilke omstændigheder.

### ***BMDL hørende til BMD fundet ved model-gennemsnit***

En del metodestudier har fokuseret på model-gennemsnit for BMDL-estimer. Nuværende praksis for at udlede en BMDL for et BMD fundet ved model-gennemsnit inkluderer beregning af et simpelt gennemsnit af BMDL-estimer fra de individuelle modeller (Kang, Kodell, and Chen 2000; Ringblom et al. 2017; Wheeler and Bailer 2009a); at bruge en metode der bygger på en varians-ulighed foreslået af Buckland et al. (Buckland, Burnham, and Augustin 1997; Namata et al. 2008; Piegorsch et al. 2013); eller ved hjælp af bootstrap (Buckley, Piegorsch, and West 2009; Roberts and Martin 2010; Wheeler and Bailer 2007). Den førstnævnte tilgang er den enkleste, men den har ingen dybere teoretisk begrundelse. Den anden fremgangsmåde er enkel at beregne, men den resulterer i alt for lave BMDL-estimer, da korrelationer mellem

de individuelle estimater ignoreres (Piegorisch et al. 2013). En asymptotisk metode til at estimere og tage højde for korrelationer mellem BMD-estimer i BMDL fundet ved model-gennemsnit blev foreslået af Jensen og Ritz (Jensen and Ritz 2015).

I mange anvendelser kan bootstrapping betragtes som state-of-the-art på grund af dens velkendte evne til at kvantificere komplekse fordelinger på baggrund af få eller ingen antagelser (Davison and Hinkley 2007). Forskellige bootstrap metoder er blevet foreslået (Faes et al. 2007; Moon et al. 2013; Wheeler and Bailer 2007). Faes et al. (Faes et al. 2007) foreslog at estimere variansen for BMD fundet ved model-gennemsnit gennem en udvidelse af variansformlen af Buckland et al. (Buckland, Burnham, and Augustin 1997) ved at tage højde for korrelationer mellem modeller ved hjælp af bootstrapping. Moon et al. (Moon et al. 2013) foreslog også en variansformel og forskellige bootstrapping-strategier, herunder en ikke-parametrisk bootstrap-tilgang til estimering af variansen og en bias-korrigeret bootstrap-tilgang til at estimere percentilen. Kim et al. (Kim, Kodell, and Moon 2014) foreslog en lignende, men mere generel, variansformel og benyttede en bootstrapping-strategi med genbrug af bootstrap-estimer til at estimere korrelationer. Bootstrap-metoder har dog vist sig at være ustabile når de bruges på små datasæt og når en stor del af responsværdierne er 0 (ingen hændelser) (Buckley, Piegorisch, and West 2009). Desuden fandt Wheeler og Bailer (Wheeler and Bailer 2007) bias i positiv retning og konkluderede at percentil-baseret bootstrap-konfidensintervaller muligvis ikke er passende.

Simmons et al. (Simmons et al. 2015) foreslog en fuld bayesiask tilgang der anvender a posteriori model-sandsynligheder som vægte til modelgennemsnit og en simulationsstrategi for at finde a posteriori fordelingen for BMD-estimatet. BMDL-estimatet bestemmes som 5% percentilen fra denne fordeling. Shao og Small (Shao and Small 2011) anvendte en Markov Chain Monte Carlo-strategi ved at generere nye værdier for parametrene i dosis-respons-modellen baseret på startværdier og en fælles normalfordeling.

#### **4.2.4 Hierarkiske eksperimenter og inklusion af kovariat information**

Reproduktions-toksicitets analyser samt eksperimenter med både tekniske og biologiske gentagelser bruger som oftest et design med en hierarkisk eller nestet struktur. Der er dybest set fire tilgange til håndtering af sådanne data: 1) Ritz et al. (Ritz, Gerhard, and Hothorn 2013) foreslog en generel fremgangsmåde til at estimere BMD og BMDL ved at bruge ikke-lineære mixed effects modeller til kontinuerte data kombineret med en modificeret version af hybridmetoden. 2) Zhu et al. (Zhu, Wang, and Jelsovsky 2007) foreslog en generaliseret estimeringsligning (GEE) metode til håndtering af grupperede binomiale respons-værdier ved at bruge udvidede generaliserede lineære modeller med en parameterafhængig linkfunktion. 3) Generaliserede lineære mixed effects modeller (Mcculloch and Neuhaus 2001) kan bruges hvis nedre og øvre grænse kan antages af være henholdsvis 0 og 1. 4) Zhang et al. (Zhang, Bailer, and Oris 2014) foreslog en bayesiask metode til at håndtere hierarkiske responsdata. For eksempel for tælldata med mange nuller fra reproduktions-toksicitetstest, hvor dødeligheden af organismerne også er dosisafhængig (Zhang, Bailer, and Oris 2012b), eller for hierarkiske binære respons data, hvor modellen for hver af svarene vil være betinget af den foregående progression, for eksempel vil deformitet afhænge af overlevelse, der afhænger af udklækning (Zhang, Bailer, and Oris 2014). Indtil nu er der dog meget få anvendelser i publiceret litteratur om BMD-bestemmelse, der tager højde for hierarkisk information.

Selvom det er muligt at tage højde for kovariater (Slob 2018), for eksempel køn eller kropsvægt, findes der meget lidt arbejde med at indbygge kovariat-information i dosis-respons-modeller. Wheeler et al. (Wheeler et al. 2017) foreslog en bayesiask percentil-regressionsmodel der muliggjorde indbygning af kovariater, og Liu et al. (Liu et al. 2018) anvendte autoregressive logistiske modeller for at tillade kovariater og tage højde for geospatial autokorrelation.



Budtz-Jørgensen et al. (Budtz-Jørgensen et al. 2013) justerede for kovariater i en pooled analyse af flere kohorte studier ved at justere respons niveauet enten på tværs af studier eller som studie-specifikke effekter.

#### 4.2.5 Kombination af flere datasæt og integration af historiske data

Generelt kræver det at kombinere datasæt til BMD-bestemmelse biologisk vurdering og passende statistiske metoder for at redegøre for variationen mellem datasættene. Ser man bort fra variationen mellem datasæt vil det resultere i forkerte BMDL-estimerer der undervurderer usikkerheden i data. Dog har der ikke været mange studier, der fokuserer på denne udfordring (US Environmental Protection Agency 2012). Nogle få har undersøgt hvordan man kombinerer data til BMD-bestemmelse: Wheeler og Bailer (Wheeler and Bailer 2009b) kombinerede replikerede reproduktionsdata fra forskellige laboratorier der testede det samme toksiske middel ved hjælp af en hierarkisk model og MCMC-algoritmer til estimering af BMD og BMDL. I et epidemiologisk studie af Budtz-Jørgensen m.fl. blev data fra syv kohorte studier kombineret i en model der tillod studiespecifikke niveauer af responsvariablen. Dette blev gjort ved i form af et studie-specifikt intercept, ved yderligere at tillade studie-specifik residual-variation og/eller kovariat effekter, eller i form af en mixed model med studie-specifik tilfældig intercept eksponerings effekt (Budtz-Jørgensen et al. 2013). Slob (Slob 2014a) foreslog at kombinere beslægtede datasæt der kigger på samme respons variable men er forskellige med hensyn til arter, køn, eksponeringstid eller endda kemikalier ved hjælp af en kombineret model med ens eller undergruppe-afhængige parametre. Hvis nogle parametre er de samme på tværs af undergrupper, kan de resterende parametre estimeres med en højere præcision, hvilket betyder smallere konfidensintervaller for den estimerede BMD og dermed et højere BMDL der ligger tættere på BMD. For eksempel analyserede Slob og Setzer (Slob and Setzer 2014) et stort antal historiske dosis-respons datasæt og fandt at dosis-respons fra in vitro studier var ret ens med hensyn til hældningsparameteren, og Hughes et al. (Hughes, Bailer, and Denton 2001) re-analyserede et stort antal tests på spildevand og fandt lignende hældningsestimerer på tværs af forskellige arter for flere af de toksiske stoffer.

Historiske data kan integreres i BMD-estimering på flere måder for at reducere usikkerheden forbundet med BMD-estimatet. Men hvis de historiske data er meget forskellige fra de nye data, for eksempel når fordelingen af responsen varierer over tid, er der ingen garanti for en reduktion i usikkerheden (Shao 2012). Der er ikke enighed om den bedste tilgang til inkorporering af historiske data. Flere strategier diskuteres i litteraturen. En måde er at bruge historiske data til at definere baggrundssandsynligheden for kontinuerte resultater. Det kan for eksempel resultere i en reduktion af usikkerheden ved den estimerede standard afvigelsen, som ellers ville nedsætte præcisionen af den estimerede BMD for kontinuerte data når hybridmetoden anvendes (Gaylor and Chen 1996; Zeller, Duran-Pacheco, and Guérard 2017). Shao (Shao 2012) undersøgte tre forskellige fremgangsmåder til at integrere historiske data ved BMD bestemmelse baseret på bayesianske model-gennemsnit; en strategi hvor data pooler ukritisk og to bayesianske modeller. De sidstnævnte to fremgangsmåder, som integrerede de historiske data via a priori antagelser om parametre for dosis-respons-modellen, fandt ingen reel indflydelse på den resulterende estimerede BMD. Alternativt kan dosisresponsmodellen reparametriseres så BMD bliver en parameter i modellen, for hvilken der kan inkluderes en informativ prior baseret på historiske data (Fang et al. 2015; Fang, Piegorsch, and Barnes 2015). Fox et al. (Fox, Hogan, and Davis 2017) foreslog anvendelse af historiske data til at estimere den designeffekt, der blev anvendt til justering i aktuelle summary-data fra udviklingsundersøgelser, der ikke rapporterer kuldspecifikke data. En konceptuel anden metode kunne være at kombinere alle data ved hjælp af en metaanalyse-tilgang med et sæt vægte der afspejler kvaliteten af de historiske data (Ritz et al. 2019).

#### 4.2.6 Simultan inferens

Man kan estimere BMDL for forskellige BMR, for eksempel 0,01, 0,05 og 0,10, eller karakterisere formen for hvordan BMDL-estimer ændrer sig med ændrede niveauer af BMR. Når flere BMR-niveauer betragtes samtidig opretholdes det samlede nominelle niveau for konfidensintervallerne ikke længere, dvs. sandsynligheden for, at mindst en af de sande BMD ligger under den tilsvarende estimerede BMDL, er højere end det påståede nominelle niveau (Wu et al. 2006). For at imødekomme denne problemstilling kan BMDL-estimer justeres for simultan inferens, dvs. problematikken med at kigge på flere statistiske tests samtidig og hvad det betyder for risikoen for falske positive resultater. Ved en sådan justering gælder det at jo flere BMR-niveauer, jo bredere er de justerede konfidensintervaller (lavere BMDL-estimer) for de estimerede BMD-værdier. En simpel tilgang kunne være at anvende en Bonferroni-justering af de percentiler der bruges til at konstruere konfidensintervaller og i overensstemmelse hermed BMDL-estimerne. Denne justering er imidlertid velkendt for at være konservativ i tilfælde af korrelerede data, hvorfor Bonferroni-justerede BMDL-estimer normalt blive for små selv ved kun et par BMR-niveauer da de er baseret på den samme kurve og endda koncentreret på en lille del af kurven.

De fleste forsøg på at tage højde for simultan inferens er baseret på simultane konfidensbånd. Nitcheva et al. (Nitcheva et al. 2005) præsenterede en justering ved evaluering af flere BMR-niveauer (2 til 5 niveauer) estimeret ud fra binomiale respons data; deres justering resulterede i 95%-konfidensintervaller der indeholder den ukendte værdi langt mere end 95% af tiden. Al-Saidy et al. (Al-Saidy et al. 2003) foreslog konfidensbånd for multistage modellen for binomiale respons data og Buckley et al. (Buckley, Piegorsch, and West 2009) sammenlignede denne fremgangsmåde anvendt på ettrinsmodellen med et profil-likelihood konfidensbånd og tre forskellige bootstrap-metoder. Deres simulationer viste, at profil-likelihood metoden var bedst. Deutsch og Piegorsch (Deutsch and Piegorsch 2013) foreslog et konfidensbånd ved at kombinere simultane Scheffé-konfidensområder med delta-metoden. Deres simulationer viste, at deres bud på konfidensbånd resulterede i ekstremt konservative BMDL-estimer. Lignende resultater rapporterede Buckley og Piegorsch (Buckley and Piegorsch 2008) der arbejdede med Abbott-justerede log-logistiske og Weibull modeller. Peng et al. (Peng, Robichaud, and Alsubie 2015) overvejede tre forskellige typer konfidensbånd for multi-stage modellen; et ensidet hyperbolisk bånd, et konstant-bredde bånd og et bootstrap-konfidensbånd (West, Nitcheva, and Piegorsch 2009). Deres resultater viste at alle tre metoder resulterede i ekstremt konservative dækningsandsynligheder. Kerns (Kerns 2017, 2018) sammenlignede simultane hyperboliske og tre-segmenters konfidensbånd for den Abbot-justerede log-logistiske model og fandt ligeledes konservative konfidensintervaller.

Det kan generelt forventes at konfidensbånd resulterer i konservative BMDL estimer da de er lavet til simultan inferens for hele dosisområdet og ikke kun de potentielt få doser af interesse. Jensen og Ritz (Jensen and Ritz 2015) foreslog at finde simultane BMDL-estimer som nedre grænse af Wald-konfidensintervaller baseret på sandwich-kovarians estimer med justerede kritiske værdier, der tager hensyn til korrelationen mellem de estimerede BMD-værdier, hvilket svarer til en tilnærmelse af konfidensområder med multi-dimensionelle kasser.

Et andet relateret problem er, hvordan man inkorporerer mere end en responsvariabel til BMD-bestemmelse. For eksempel er dette tilfældet i vand-toksikologiske undersøgelser, hvor fokus er på klækning, overlevelse og vækst af afkom. Det giver en hierarkisk struktur med 2 binomiale respons variable der går forud for en kontinuert respons variable (Zhang, Bailer, and Oris 2012b, 2014). Når hver respons variabel betragtes separat, er en simpel tilgang blot at rapportere det laveste BMDL som det mest følsomme respons (Hardy et al. 2017). Denne tilgang er imidlertid ikke i overensstemmelse med definitionen af BMDL. Budtz-Jørgensen (Budtz-Jørgensen 2007) foreslog en strukturel ligningsmodel der kombinerer effekten af en flerdimensionel eksponeringsvariable på et flerdimensionelt respons efter justering for målefejl, for på den måde at udnytte sammenhængen mellem de forskellige respons variable mere optimalt. Yu og Catalano (Yu and Catalano 2005) foreslog en metode der kombinerer risikoen fra to

kontinuerte respons variable. Her blev BMD defineret, med en metode, der lignede hybridtilgangen, dvs. den dosis hvor en uønsket event blev observeret for mindst et af de to udfald. Metoden fra Jensen og Ritz (Jensen and Ritz 2015)(85) muliggjorde simultan inferens på tværs af forskellige respons variable af forskellig type (f.eks. binomiale og kontinuerte). Najita et al. (Najita, Li, and Catalano 2009) tilpassede en likelihood-baseret metode til fælles modellering af en binomial og en kontinuert respons variable, hvor det kontinuerte respons mangler som en del af designet afhængigt af resultatet af det binomiale respons. Endelig skal det nævnes at Deutsch og Piegorsch (Deutsch and Piegorsch 2012, 2013) introducerede BMD-profiler i to dimensioner.

#### 4.2.7 Retningslinjer og regulativer

I 2009 offentliggjorde EFSA et vejledningsdokument til brugen af BMD-metoden i risikovurdering (European Food Safety Authority scientific committee 2009), som senere er blevet opdateret (Hardy et al. 2017). Med opdateringen er der en forskydning af præference fra at bruge den laveste BMDL af flere modeller til modeludvælgelse baseret på AIC og visuel vurdering eller modelgennemsnit. Sidstnævnte fremhæves også i den seneste litteratur på grund af dens robusthed. EFSA-dokumentet forklarer BMD-konceptet i detaljer med eksempler; der foreslår et rapporteringsformat og peger i retning af mulige forbedringer af det eksperimentelle design. EFSA foreslår også at bruge software, nemlig 'BMDS' (Davis, Gift, and Zhao 2011) og 'PROAST' (Slob 2018).

I deres tekniske vejledning fra 2012 (US Environmental Protection Agency 2012) anbefaler US EPA at man bruger BMD-metodikken frem for NOAEL / LOAEL-metoden for at harmonisere metoder til kræft og ikke-kræft risikovurdering. Retningslinjerne fra US EPA fra 2012 anbefaler imidlertid endnu ikke brug af modelgennemsnit. Der er kun begrænset vejledning tilgængelig fra OECD og kun med hensyn til studier på kroniske og kræftfremkaldende effekter (OECD 2014). Yderligere aspekter er blevet diskuteret af Haber et al. (Haber et al. 2018) med særlig fokus på forskelle mellem US EPA og EFSA. Derudover viser man, hvordan forskellige resultater kan nås gennem forskellige fortolkninger af de eksisterende retningslinjer.

I modsætning til den tilsyneladende udbredte anerkendelse af BMD-metoden fra internationale og nationale institutioner og myndigheder, er der nogle begrænsninger med hensyn til harmonisering. Nogle regulativer kræver specifikt rapportering af NOAEL, såsom regulativet om datakrav til aktive stoffer til brug i plantebeskyttelsesmidler (Reg. (EU) 283/2013), og henviser kun til muligheden for at udlede "om nødvendigt, andet passende reference værdi" og her specifikt kun for langtids / kræftisiko-studier. Dette svarer også til de respektive test-guidelines (TG): OECD-TG 451 (kræftfremkaldende virkning) (Organisation for Economic Co-operation and Development 2009a), 452 (kronisk toksicitet) (Organisation for Economic Co-operation and Development 2009b) og 453 (kombineret kræftfremkaldende virkning og kronisk toksicitet)(Organisation for Economic Co-operation and Development 2009c), som implicit dækker BMD. På lignende måde gør de nyreviderede regulativer for toksicitetsstudier på gentagen dosering såsom: OECD TG 412 (Organisation for Economic Co-operation and Development 2017a) og 413 (Organisation for Economic Co-operation and Development 2017b) (inhalationstoksicitet ved gentagen dosis). Nogle testretningslinjer for andre respons variable mangler dog stadig henvisningen til BMD, for eksempel 407 (28-dages oral toksicitet) (Organisation for Economic Co-operation and Development 2008), 408/409 (90-dages gentagen oral toksicitet)(Organisation for Economic Co-operation and Development 1998a, 1998b) eller 416 (to-generation reprotoksicitet) (Organisation for Economic Co-operation and Development 2001). I nogle studier, f.eks. genotoksicitetsstudier accepteres BMD-fremgangsmåden endnu ikke men det kan ændre sig i fremtiden (MacGregor et al. 2015).

Harmonisering af testretningslinjerne er nødvendig for at muliggøre en globalt harmoniseret regulativ vurdering. En øget statistisk styrke kan ofte opnås ved at ændre på gruppestørrelser uden at ændre det samlede antal dyr. Mens en sådan strategi anerkendes af EFSA (Hardy et al. 2017) og dermed kan betragtes som værende en holdbar løsning i Europa, kan det imidlertid utilsigtet øge antallet af dyr i forsøg til globale registreringsformål. Et andet nødvendigt trin

på vejen til en mere udbredt regulatorisk anvendelse er en diskussion af hvordan man håndterer mindre informative data, dvs. data, der resulterer i meget brede konfidensintervaller for BMD og altså BMDL-værdier der ligger langt fra BMD, fordi dette fra et statistisk perspektiv kan udløse et behov for at generere nye data, som dog kan være underlagt dyrevelfærd eller andre specifikke regulatoriske overvejelser.

### 4.3 Opsummering

I løbet af de sidste to årtier har BMD-metoden fået øget opmærksomhed, og antallet af anvendelser er øget drastisk. BMD bruges nu også uden for dets oprindelige tiltænkte anvendelser. Parallelt med stigningen i anvendelser har der været en omfattende udvikling i BMD-metoder med en over-repræsentation af metoder for binomiale i forhold til kontinuerte respons data. På trods af den øgede brug af BMD og den løbende udvikling af metoden er der stadig nogle udfordringer, især relateret til software, BMDL-estimering og kontinuerte responsdata, hvor både BMR såvel som baggrundsniveaue,  $p_0$ , afhænger af den skala, hvorpå responsen måles.

Selvom der er sket fremskridt inden for BMD-metoderne, har vi identificeret en række områder, hvor der er presserende behov for yderligere arbejde. Især fandt vi:

1. mangel på fleksibel og brugervenlig software,
2. ingen konsensus om, hvordan man kan udlede BMDL (hvilke konfidensintervaller skal der bruges?),
3. meget konservativ justering, hvis flere BMR-værdier overvejes (problemerne inden for simultan inferens)
4. ingen definition af BMD for event-time-data
5. ingen generel modelleringsmetode til mere komplekse eksperimentelle design.

Der mangler desuden grundige sammenligninger af forskellige metoder inden for BMD-metodologien.

# 5. *bmd* – en R-pakke til bestemmelse af BMD

## 5.1 Baggrund

### 5.1.1 Eksisterende løsninger

Der er flere eksisterende løsninger til bestemmelse af BMD, hvor de to primært dominerende software programmer er BMDS udviklet af U.S. EPA (Davis, Gift, and Zhao 2011) og PROAST udviklet til EFSA i et samarbejde mellem det Hollandske nationale institut for folkesundhed og miljø (RIVM) og Open Analytics (Slob 2018; Varewyck and Verbeke 2017).

De to programmer adskiller sig på flere måder. En sammenligning på adskillige parametre kan findes i Jensen, Kluxen, and Ritz (2019) samt i tabel 1. Kort opsummeret kan det nævnes at BMDS, der er et Windows-shell-program, tilbyder flest dosis-respons-modeller og forskellige definitioner af BMD (beskrevet tidligere), men der mangler (for nu) model-gennemsnit.

PROAST, der er udviklet i det statistiske software **R** men kører som et separat program, tilbyder færre dosis-respons-modeller og kun en enkelt definition ("excess risk") for binomiale data og en (relativ ændring) for kontinuerte data, men giver til gengæld mulighed for model-gennemsnit for både binære og kontinuerte responsdata. Begge programmer estimerer BMDL ved hjælp af profil-likelihood tilgangen; dog tilbyder PROAST også bootstrap metoden.

Af andre eksisterende programmer til BMD bestemmelse kan nævnes MADr-BMD (Wheeler and Bailer 2008) der er et kommandoprompt program udviklet til EFSA med det primære formål at håndtere model-gennemsnit, dog kun for binære endpoints, og BMDExpress (Yang, Allen, and Thomas 2007) der er et Windows-shell-program udviklet til bestemmelse af BMD for high-throughput data.

Der har været adskillige forsøg på at inkorporere BMD-estimering for high-throughput omics-data, nogle med tilsvarende softwareudvikling (Burgoon and Zacharewski 2008; Smetanová et al. 2015; Thomas et al. 2007; Yang, Allen, and Thomas 2007). Der synes dog stadig at være en del udfordringer med hensyn til modeludvælgelse og modelkontrol (Smetanová et al. 2015; Yang, Allen, and Thomas 2007), valg af BMR for kontinuerte data (Thomas et al. 2007) og hvordan man definerer og kombinerer gruppen af gener der bruges til at udlede den estimerede BMD (Farmahin et al. 2017; Thomas et al. 2011).

**Table 1:** Oversigt over funktionaliteter indbygget i bmd-pakken sammenlignet med de tre nuværende mest brugte software programmer til BMD bestemmelse.

		BMDS	PROAST	MADr-BMD	bmd
Tillader kuld-effekter		X	X		(X) <sup>1</sup>
Tillader kovariater			(X) <sup>2</sup>		(X) <sup>2</sup>
Definitioner for binomiale respons data	Added risk	X	X	X	X
	Excess risk	X	X	X	X
	Direkte specification				X
Definitioner for kontinuerte respons data	Hybrid metoden, additional risk	X			X
	Hybrid metoden, excess risk				X
	“added response”				X
	“relative change”	X	X		X
	“extra response”	(X) <sup>3</sup>			X
	Direkte definition	X			X
Definitioner for tælle respons data	“added response”				X
	“relative change”				X
	“extra response”				X
	Direkte definition				X
BMDL estimation	Delta metoden				
	Invers regression				
	Profil likelihood	X	X		
	Bootstrap		X	X <sup>4</sup>	X <sup>5</sup>
Model-gennemsnit for binomiale respons data			X	X	X
Model-gennemsnit for kontinuerte respons data			X		X
Model-gennemsnit for tælle respons data					X

<sup>1</sup> Kan håndteres med sandwich estimator

<sup>2</sup> Der stratificeres på kovariat niveau i PROAST. Det kan man naturligvis også i bmd-pakken

<sup>3</sup> Kun for Hill modellen

<sup>4</sup> Parametrisk bootstrap

<sup>5</sup> Ikke-parametrisk, semi-parametrisk eller parametrisk bootstrap. Semi-parametrisk kun for kontinuerte respons data, mens semi-parametrisk kan bruges for kontinuerte og binomiale respons data.

## 5.1.2 Problemstilling

De eksisterende løsninger har alle deres styrker men også svagheder. Derfor kan valg af software afhænge af de data, der skal analyseres. Som det ser ud nu er det kun den statistiske software R, der muliggør anvendelse af alle aktuelle metoder til BMD og BMDL bestemmelse. Dog kan anvendeligheden diskuteres, da **R** er programmerings-baseret og meget få af de tilgængelige metoder er implementeret (hvis vi her ser bort fra *bmd*-pakken), hvilket efterlader en vis programmering for slutbrugeren.

Et stort problem er at de fleste programmer til bestemmelse af BMD stadig er begrænsede med hensyn til de implementerede dosis-respons-modeller. Det fulde potentiale ved dosis-respons-modellering, som for eksempel demonstreret af R-pakken *drc* (Ritz et al. 2019), er ikke fuldt ud realiseret i forbindelse med BMD-analyse. For eksempel er der – efter vores bedste viden - med de tilgængelige værktøjer ikke mulighed for at bestemme BMD og BMDL for tælle-data selvom koncepterne vil kunne anvendes på samme måde for denne type dosis-respons-data. Med de tilgængelige BMD-software programmer mangler desuden en mulighed for at håndtere situationer hvor antallet af parametre i dosis-respons-modellerne er større end antallet af doser (Slob 2014a). Både EFSA og den amerikanske EPA-vejledning kræver specifik BMD-uddannelse og mere raffineret software til at udnytte nye statistiske metoder.

## 5.2 R-pakken

For at imødekomme flere af problemstillinger fremstillet ovenfor, har omdrejningspunktet i dette projekt været, at lave en **R**-pakke specifikt designet til at bestemme BMD. I løbet af de sidste 25 år har open source-miljøet **R** udviklet sig til et stærkt og alsidigt statistisk software. Udviklingen af en stor mangfoldighed af nyttige udvidelsespakker gør systemet nyttigt inden for en række videnskabelige områder. Blandt disse pakker er *drc* specielt udviklet til analyse af dosis-responsdata. Med den brugervenlige grænseflade til specifikation af modelantagelser om dosis-respons-sammenhængen, det fleksible sæt indbyggede modelfunktioner inklusive selvstarterfunktioner og lethedens hvormed man kan finde afledte parameter er *drc* blevet et velkendt og meget anvendt valg til dosis-respons-analyse. R-pakken *bmd* der er udviklet i dette projekt bygger på fleksibiliteten i *drc*, hvorfor den faciliterer BMD og BMDL bestemmelse for en meget bred vifte af dosis-respons-modeller. I *bmd*-pakken er der endvidere indbygget nogle af de nyeste metoder til BMD bestemmelse.

### 5.2.1 Nøglefunktionen *bmd()*

R-pakken er en samling af flere funktioner. Nøgle funktionen er *bmd()* der tager følgende argumenter, hvoraf de første fire skal specificeres:

- *object* – Et model fit fra *drc*
- *bmr* – BMR niveauet for hvilket BMD skal beregnes
- *backgType* - Hvordan skal baggrundsniveauet specificeres
- *def* – Definitionen der skal bruges til at beregne BMD
- *backg* – Argument til specifikation af baggrundsniveauet hvis dette er brugerdefineret
- *controlSD* – Specifikation af antal standard deviationer der skal bruges til at definere en ad-verse hændelse, hvis BMD er defineret ved hybrid metoden
- *interval* - Metode til beregning af BMDL
- *sandwich.vcov* - Specificerer om BMDL skal beregnes med brug af robuste standardfejl

### 5.2.2 Definitioner af BMD

Forskellige typer af ændringer i respons variabelen kan være relevante, det være sig absolutte eller relative ændringer, hvorfor der er forskellige definitioner tilstede i *bmd*-pakken. Ikke alle definitioner er relevante for alle typer af respons data. I tabel 2 ses en opsummering af hvilke definitioner der er til rådighed for de forskellige typer af respons data.

**Tabel 2:** Oversigt over de forskellige definitioner af BMD tilgængelige i bmd pakken (med argumentet def). Et X indikerer at definitionen er relevant for denne type respons data

Definition	def	Binomiale data	Kontinuerte data	Tælle data
Additional	“additional”	X		
Excess/extra	“excess”	X		
Hybrid approach (Additional risk)	“hybridAdd”		X	
Hybrid approach (Excess risk)	“hybridExc”		X	
Absolut ændring	“added”		X	X
Relativ ændring	“relative”		X	X
Extra respons	“extra”		X	X
Direkte defineret	“point”	X	X	X

### 5.2.3 Bestemmelse af baggrunds niveauet

Da BMD er defineret som den dosis, der resulterer i en prædefineret ændring i respons i forhold til baggrunds niveauet, kan det have stor betydning hvordan baggrunds niveauet findes. I *bmd* kan baggrunds niveauet estimeres fra dosis-respons-modellen (`backgType="modelBased"`) eller brugeren kan specificere det direkte (`backgType="absolute"`) gennem argumentet `backg`. Hvis BMD ønskes bestemt ved hjælp af hybrid metoden, kan man vælge om baggrunds niveauet skal defineres via en percentil i fordelingen for kontrolgruppen eller i termer af standard afvigelser fra middelværdien i kontrol gruppen. I sidstnævnte tilfælde er der mulighed for at lade standard afvigelsen være baseret på modellen eller selv at beregne og angive den (gennem argumentet `controlSD`).

Der kan være flere grunde til selv at ville angive enten baggrunds niveau eller kontrol standard afvigelse. Herunder kan argumentet om brug af historiske kontrol data veje tungt.

### 5.2.4 Beregning af BMDL

Den nedre grænse af et en-sidet konfidensinterval for BMD, BMDL, kan bestemmes på fire måder i *bmd*-pakken. Deafult er at bruge standard Wald konfidensintervaller baseret på delta-metoden. En anden mulighed er at bruge invers regression, hvor BMDL findes som den dosis der resulterer i at øvre grænse for konfidensbåndet omkring dosis-respons kurven rammer den prædefinerede ændring i responset. Denne metode kræver dog ofte meget data eller at der ses på store ændringer (bruges et højt BMR).

En tredje mulighed for at bestemme BMDL opnås ved at kombinere delta metoden med sandwich-variens-covarians estimator. Når antagelser om normalitet og / eller varianshomogenitet ikke er opfyldt, kan der opnås konsistente estimater af standardfejlene ved at justere den estimerede variens-covariansmatrix. De resulterende justerede standardfejl benævnes sandwich estimator.

Som en sidste mulighed kan BMDL også findes ved bootstrapping. I *bmd*-pakken er der mulighed for både ikke-parametrisk (default), semi-parametrisk og parametrisk bootstrapping alt afhængigt af hvilken type respons der analyseres. Bootstrap bestemt BMDL opnås ved hjælp af funktionen `bmdBoot()` der tager de samme argumenter som `bmd()`. Desuden kan man specificere antallet af bootstrap samples samt om BMDL skal beregnes ved hjælp af et percentil interval eller ved en bias korrigerede (BCa) intervaller.

### 5.2.5 Bestemmelse af BMD ved model-gennemsnit

For at bestemme BMD og BMDL ved hjælp af model-gennemsnit, bruger man funktionen `bmdMA()`, der for de fleste parametres vedkommende følger `bmd()`. Dog er første argument nu en liste af model fits (en liste af kandidat modeller).

To forskellige fremgangsmåder kan anvendes til bestemmelse af BMD ved model-gennemsnit. En fremgangsmåde er at lave et vægtet gennemsnit af BMD-estimerne fra alle kandidatmo-



dellerne. Den anden er at lave et vægtet gennemsnit af de fittede kurver fra alle kandidatmodellerne og derefter finde BMD som den dosis der resulterer i ændringen af interesse i responsen prædikeret ved denne vægtede model (Wheeler and Bailer 2007, 2008). Vægte kan enten specificeres af brugeren eller defineres ved hjælp af AIC (default) eller BIC, to velkendte og ofte benyttede informationskriterier.

I `bmdMA()` er det indbygget tre forskellige metoder til beregning af BMDL ved argumentet `type`. Den første mulighed er at bruge Wald-konfidensintervaller hvor standardfejlen er approksimeret ved en variansulighed af Buckland et al. (1997). Den anden mulighed er at bruge et vægtet gennemsnit af BMDL fra alle kandidatmodellerne, på samme måde som BMD model-gennemsnittet beregnes (Kang, Kodell, and Chen 2000). Endelig kan BMDL findes ved bootstrapping.

### 5.3 Opsummering

Den udviklede R-pakke imødekommer nogle af de udfordringer relateret til estimation af BMD og BMDL, der ikke imødekommes af BMDS og PROAST samt samler mulighederne der er at finde i de to software løsninger så man som bruger kan nøjes med at lave alle sine analyser et sted. Da R-pakken `drc` i forvejen er et populært valg til dosis-respons analyser bliver den nye R-pakke en simpel udbygning af data analysen i et og samme software system.

R-pakken `bmd` udbygger mulighederne og adskiller sig blandt andet fra BMDS og PROAST på følgende måder:

- Der kan bestemmes BMD for tælle data.
- Det er langt nemmere at arbejde videre med output direkte, hvilket for eksempel muliggør at der kan bestemmes BMD for tid-til-event data.
- Med undtagelse af bestemmelse af BMDL ved profil-likelihood metoden er alle funktionaliteter fra BMDS og PROAST samlet ét sted, herunder model-gennemsnit for alle typer af responsdata.
- R-pakken muliggør bestemmelse af BMD for stort set alle dosis-respons-funktioner til rådighed i BMDS og PROAST. Derudover, er der mulighed for at bruge fraktionelle-polynomier, der er særligt velegnet til model-gennemsnit, flere hormesis modeller, samt semi-parametriske alternativer.
- I den udviklede R-pakke er der desuden mulighed for at beregne BMDL baseret på sandwich-variens-covarians estimator, der giver robuste standardfejl og dermed gør det muligt at tage højde for model misspecifikationer relateret til fordelingsantagelserne.
- BMDL kan bestemmes ved flere forskellige former for bootstrap.
- Der er flere muligheder for bestemmelse af BMD og ikke mindst BMDL gennem model-gennemsnit.

# 6. Eksempler på brug af BMD i økotoxikologi

Dette afsnit indeholder syv eksempler på hvordan BMD kan bruges inden for økotoxikologi. Alle beregninger af BMD og BMDL i disse eksempler er foretaget med R-pakken *bmd*. R-kode til alle eksempler samt datasæt kan findes på GitHub under SigneMJensen/MST.

## 6.1 Binomiale data: Regnorme test på chloroacetamide

### 6.1.1 Problemstilling

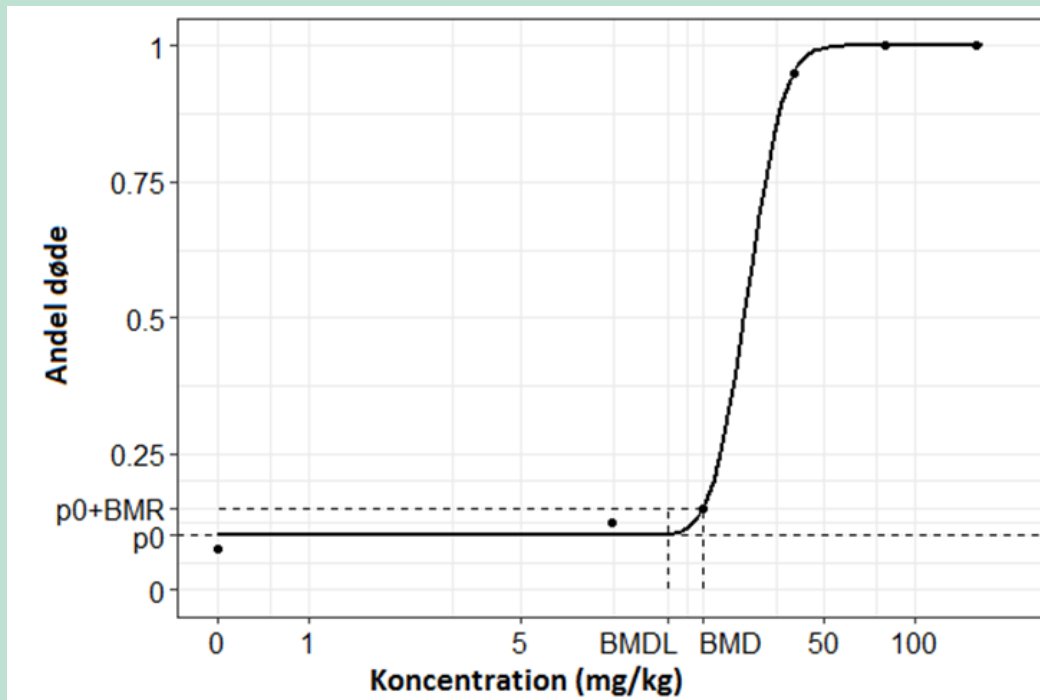
Dette eksempel illustrerer BMD i dets helt simple og oprindeligt tænkte udgave for binomiale respons data hvor BMD kan defineres som dosis resulterende i en absolut ændring ("additional risk" definitionen) i andelen af døde i forhold til kontrol populationen eller en relativ ændring ("excess risk" definitionen) i andelen af døde i forhold til kontrol populationen.

### 6.1.2 Data

Vores første eksempel gennemgår data fra et regnorme toksicitetstest (Hoekstra 1987). For hver af seks koncentrationer (fem aktive og en kontrol) blev 40 regnorme udsat for herbicidet chloroacetamid og det resulterende antal døde regnorm blev talt. Dødeligheden for kontrolgruppen, den naturlige dødelighed, var 3/40.

### 6.1.3 Analyse af data

Data blev analyseret med en tre-parameter log-normal model, som inkluderer en nedre grænse der fanger den naturlige dødelighed. Figur 5 viser den resulterende model. Den modelbaserede estimerede naturlige dødelighed var 0,10 (95% CI: 0,03-0,17).



**FIGUR 5.** Den estimerede koncentrations-responskurve for en tre-parameter log-normal model for data fra regnorme toksicitetstesten vist sammen med data (middelværdier pr. dosis) og den estimerede BMD og BMDL bestemt ved "additional risk" definitionen.

BMD blev fundet ved at bruge "additional risk" definitionen med BMR på 5% og var den dosis, der var associeret med en andel af døde regnorme lig med

$$\text{BMR} + p_0 = 0,10 + 0,05 = 0,15.$$

De estimerede BMD og BMDL-værdier var henholdsvis 20,02 mg/kg og 15,40 mg/kg.

BMD, som blev fundet ved at bruge "excess risk" definitionen, var dosis resulterende i en andel af døde regnorme lig med

$$\text{BMR} \cdot (1-p_0) + p_0 = 0,05 \cdot (1 - 0,10) + 0,10 = 0,145$$

De resulterende estimerede BMD og BMDL for "excess risk" definitionen var henholdsvis 19,79 mg/kg og 15,15 mg/kg.

Bemærk at de to definitioner for binomiale respons data er ens, hvis baggrundssandsynligheden er 0.

## 6.2 Tælle data: Toksicitet test med kobber under varierende temperatur

### 6.2.1 Problemstilling

I tests hvor der ses på reproduktion kan det give problemer/ fejlslutninger hvis antal afkom modelleres med en model der ikke tager hensyn til for tidlig dødelighed. En måde at imødekomme den problemstilling på, er at anvende en vægtet dosis-respons-model der tager hensyn til levetiden for det enkelte individ (Delignette-Muller et al. 2014). Det er dog ikke en standard model antagelse der er tilgængelig i eksisterende BMD-specialiserede software programmer.

Et andet problem relateret til tælledata mere generelt er, at de ofte udviser overdispersion, dvs. mere variation end modellen kan forklare. Der er flere metoder der kan bruges til at imødekomme denne udfordring. Den underliggende fordelingsantagelse kan ændres til f.eks. en negativ binomial fordeling der indeholder en ekstra parameter der kan fange overskydende variation. En anden mulighed er at justere standardfejlene. Ingen af disse løsninger er dog mulige i de BMD-specialiserede software programmer der heller ikke indeholder muligheden for modellering af tælle data med standard antagelsen om Poisson fordelte data.

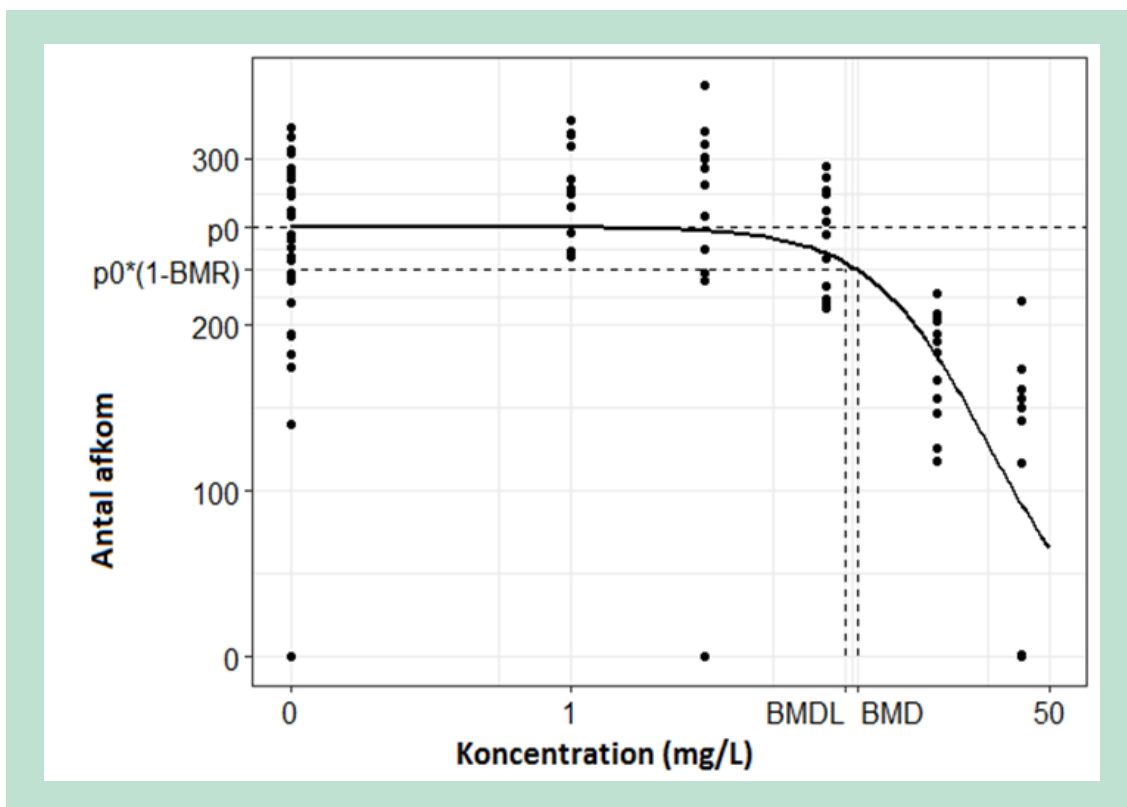
### 6.2.2 Data

Cedergreen et al. (Cedergreen et al. 2016) undersøgte effekten af varierende temperatur på toksiciteten af kobber på nematode ormen *Caenorhabditis elegans* (Maupas). Som en del af eksperimentet blev 12 nematoder udsat for hver af 5 koncentrationer af kobber (1, 3, 8, 20, 40 mg/L agar) under varierende temperatur med daglige udsving på  $\pm 4$  grader omkring en gennemsnitlig temperatur på 20 °C. Derudover blev 36 orme udsat for de samme temperatursvingninger men blev betragtet som kontroller uden eksponering for kobber. Under reproduktionsfasen af eksperimentet blev nematoderne flyttet til nye rum på friske plader hver dag. På den måde blev det muligt at registrere den daglige ægproduktion og levetiden for hver enkelt nematode. Antallet af frugtbare æg og udklækkede unge blev talt som afkom.

### 6.2.3 Analyse af data

Vi estimerede BMD for et BMR på 0,1 under anvendelse af den relative definition, dvs. koncentrationen forbundet med en 10% reduktion i det samlede antal afkom i forhold til kontrolgruppen. Der blev brugt en tre-parameter log-logistisk model med en antagelse om Poissonfordelte respons data. Figur 6 viser den resulterende model. BMD og BMDL var henholdsvis 10,47 mg/L agar og 9,40 mg/L agar.

Som et alternativ kan data modelleres med en vægtet dosis-responsmodel under hensyntagen til nematodernes levetid. BMD fra denne nye model skal nu fortolkes som koncentrationen forbundet med en 10% ændring i antallet af afkom pr. dag i forhold til kontrolgruppen. På grund af de to forskellige fortolkninger kan BMD- og BMDL-estimer fra den vægtede model (henholdsvis 31,16 mg/L og 28,89 mg/L) ikke sammenlignes direkte med dem fra den første model.



**FIGUR 6.** Den estimerede koncentrations-responskurve for en tre-parameter log-logistisk model tilpasset data fra toksicitetstesten med kobber under forskellige temperaturer. Modellen er vist sammen med alle datapunkter og den estimerede BMD og BMDL fundet ved den relative definition og BMR på 10%.

Hvis den underliggende fordelingsantagelse ændres fra Poisson til negativ binomial fordelte data bliver BMD og BMDL nu henholdsvis 27,99 mg/L og 15,99 mg/L. Den anden mulighed for at imødekomme et potentielt problem med overdispersion er at bruge sandwichestimerer når man estimerer BMDL. Det kræver dog at man stoler på at dosis-responsmodellen beskriver gennemsnitstendensen i data korrekt. Resultatet af denne fremgangsmåde var en BMD på 31,16 mg/L og en BMDL på 18,63 mg/L. Bemærk, at begge procedurer der justerer for overdispersion, resulterede i et stort fald i BMDL. Det indikerer at overdispersion faktisk var til stede i disse data.

## 6.3 Kontinuerte data og model-gennemsnit: Fiske test

### 6.3.1 Problemstilling

For kontinuerte data, er det ikke altid oplagt hvor stor en effekt der skal til før man vil betegne det som værende en uønsket effekt. Af samme årsag kan det være svært at definere hvad baggrundssandsynligheden for en uønsket effekt er. Hybrid metoden kan ses som en måde at imødekomme denne udfordring. Her defineres uønskede effekter som de værdier der ligger langt ude i halen af fordelingen af respons-værdierne for kontrolgruppen. Disse værdier er, per definition, "unormale" for kontrolgruppen, hvorfor det giver god mening at bruge et cut-off i fordelingen for kontrollerne til at definere baggrundsniveauet.

Interpolering i et dosis-område hvor der er meget få eller ingen data, som typisk er tilfældet i en BMD-analyse, er meget afhængig af model-fit og af valget af dosis-responsmodel. For delvist at overvinde dette er det blevet foreslået at evaluere flere modeller og efterfølgende vælge den bedste model bestemt ved hjælp af et eller andet kriterium for godness-of-fit (Slob 2002).

Usikkerheden involveret i modeludvælgelsesprocessen er dog ikke inkorporeret i BMD og den tilknyttede BMDL. Faktisk kan resultater baseret på "den bedste" model resultere i biased estimater af BMD og ikke-beskyttende estimater af BMDL, dvs. for BMDL estimater der er for høje og dermed ikke opfylder det, at være lavere end det "sande" BMD i 95% af gentagelser (Ringblom, Johanson, and Öberg 2014; West et al. 2012). Som konsekvens heraf anbefales modelgennemsnit nu til BMD analyse af adskillige forfattere (Faes et al. 2007; Jensen and Ritz 2015; Kang, Kodell, and Chen 2000; Namata et al. 2008; Wheeler and Bailer 2007) samt af regulatoriske myndigheder (Hardy et al. 2017). Brugen af model-gennemsnit har den konsekvens at fleksibiliteten i sættet af kandidatmodeller til en vis grad erstatter manglen på data. En anden fordel er, at det giver mulighed for en automatiseret tilgang når først kandidatmodellerne er udvalgt.

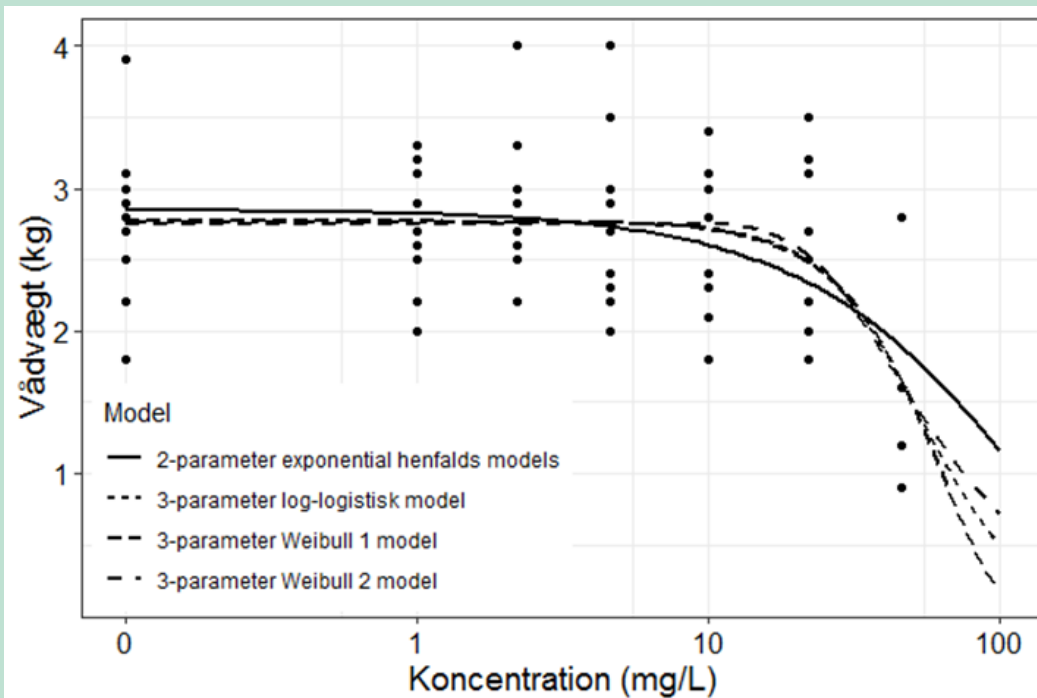
### 6.3.2 Data

Data er fra en 21 dages fisketest med Regnbueørreder, *Oncorhynchus mykiss*, brugt som test organisme og udført i overensstemmelse med retningslinjerne fra OECD GL204 (Organisation for Economic Co-operation and Development (OECD) 2006). Regnbueørrederne levede i 14-15 °C vand og blev udsat for en af 7 koncentrationer (6 + kontrol) af et ukendt middel. Efter 28 dage blev vådvægten af regnbueørrederne registreret. Der var 10 ørreder pr. koncentration. For nogle af de højere doser forekom der imidlertid manglende værdier, hvilket resulterede i et endeligt datasæt med kun 61 observationer. Ignorering af sådanne manglende værdier kan føre til biased resultater afhængigt af mekanismen, der forårsager de manglende værdier, men til illustration af BMD-metodologien vil vi ignorere det potentielle problem her.

### 6.3.3 Analyse af data

Vi fulgte OECD og fittede en to-parameter eksponentiel henfaldsmodel til data. Figur 7 (fuld optrukket linje) viser et passende fit til data, som blev understøttet af residual- og QQ-plot (ikke vist her).

For et BMR på 0,05 under anvendelse af hybridmetoden med 2 standardafvigelser som cut-off var BMD og BMDL henholdsvis 4,23 mg / L og 2,13 mg / L.



**FIGUR 7.** Den fittede koncentration-respons kurve for en to-parameter eksponentiel henfalds-model, en tre-parameter log-logistisk model samt to tre-parameter Weibull modeller fittet til regnbue ørred data vist sammen med alle data punkter.

Selvom den eksponentielle henfaldsmodel viste et rimelig fit kunne andre modeller give et lignende fit til data. For eksempel viste en tre-parameter log-logistisk model og to forskellige tre-parameter Weibull-modeller tilsvarende rimeligt visuel fit til data (se figur 7). Sammenligning af modellerne ved hjælp af AIC viste også at de fire modeller gav nogenlunde samme fit til data (tabel 3).

**Tabel 3:** BMD, BMDL, AIC, and modelvægt baseret på aIC for fire forskellige modeller fittet til data fra et fiske test med regnbue ørred. BMD blev bestemt baseret på hybrid metoden med 2 standard afvigelser som cut-off og BMR=0.05. Model-gennemsnits BMDL blev fundet ved ikke-parametrisk bootstrap.

	Parametre	BMD (mg/L)	BMDL (mg/L)	AIC	Vægt
Exponential decay	2	12.65	6.37	106.31	0.333
Log-logistic	3	22.91	9.05	106.65	0.237
Weibull type 1	3	22.72	8.12	106.58	0.253
Weibull type 2	3	24.26	14.26	106.94	0.177
Model-gennemsnit (Estimat gennemsnit)		19.68	8.04		
Model-gennemsnit (Kurve gennemsnit)		20.35	7.73		

Model-gennemsnittet baseret på de fire modeller med anvendelse af ikke-parametrisk bootstrap til bestemmelse af BMDL. Det resulterede i en meget højere BMD og en BMDL tæt på det tilsvarende estimat fra den originale (eksponentielle henfald) model, der i dette tilfælde også ville have været valgt som bedste model hvis AIC havde været brugt til at udvælge en enkelt model. Den alternative metode til at beregne model-gennemsnit baseret på hele kurver

resulterede i en lidt højere BMD, men en lidt lavere BMDL sammenlignet med model-gennemsnit beregnet på BMD estimaterne.

## 6.4 Binomiale data i et hierarkisk design: En akut toksicitet test på $\alpha$ -cypermethrin

### 6.4.1 Problemstilling

I et hierarkisk design med deleksperimenter kan en simpel dosis-respons-model muligvis ikke fange korrelationsstrukturen i data korrekt, hvorfor en model der ukritisk kombinerer data fra alle deleksperimenter uden at tage højde for strukturen i data kan lede til ubrugelige resultater. Et naturligt valg når man har at gøre med sådanne eksperimenter er at tænke i baner af mixed modeller. Mixed effekt dosis-respons-modeller kan godt estimeres for kontinuerte respons data, men for binomiale respons data er det langt mere kompliceret. En måde at komme uden om denne problemstilling er at modellere data med en to-trins metode, hvor der i første trin estimeres en dosis-respons-model for hvert uafhængigt datasæt (for eksempel deleksperiment). Andet trin kombinerer så BMD-estimerer fra hver enkelt model fra trin 1 til et enkelt estimat i en meta-analyse tilgang (Ritz et al. 2019).

### 6.4.2 Data

Gottardi og Cedergreen (Gottardi and Cedergreen 2019) undersøgte toksiciteten af  $\alpha$ -cypermethrin på *Daphnia magna* ved hjælp af en akut toksicitetstest efter OECD-retningslinjer for test af kemikalier. I ni uafhængige eksperimenter udført på forskellige tidspunkter blev nyfødte *D. magna* (<1 dag gamle) udsat for forskellige koncentrationer af  $\alpha$ -cypermethrin eller for acetone kontrol i 48 timer i M7-medium. Fire gentagelser af fem organismer blev anvendt til hver koncentration. Antallet af anvendte koncentrationer i hvert undereksperiment varierede fra 6 til 7 (5 eller 6 samt kontrol).

### 6.4.3 Analyse af data

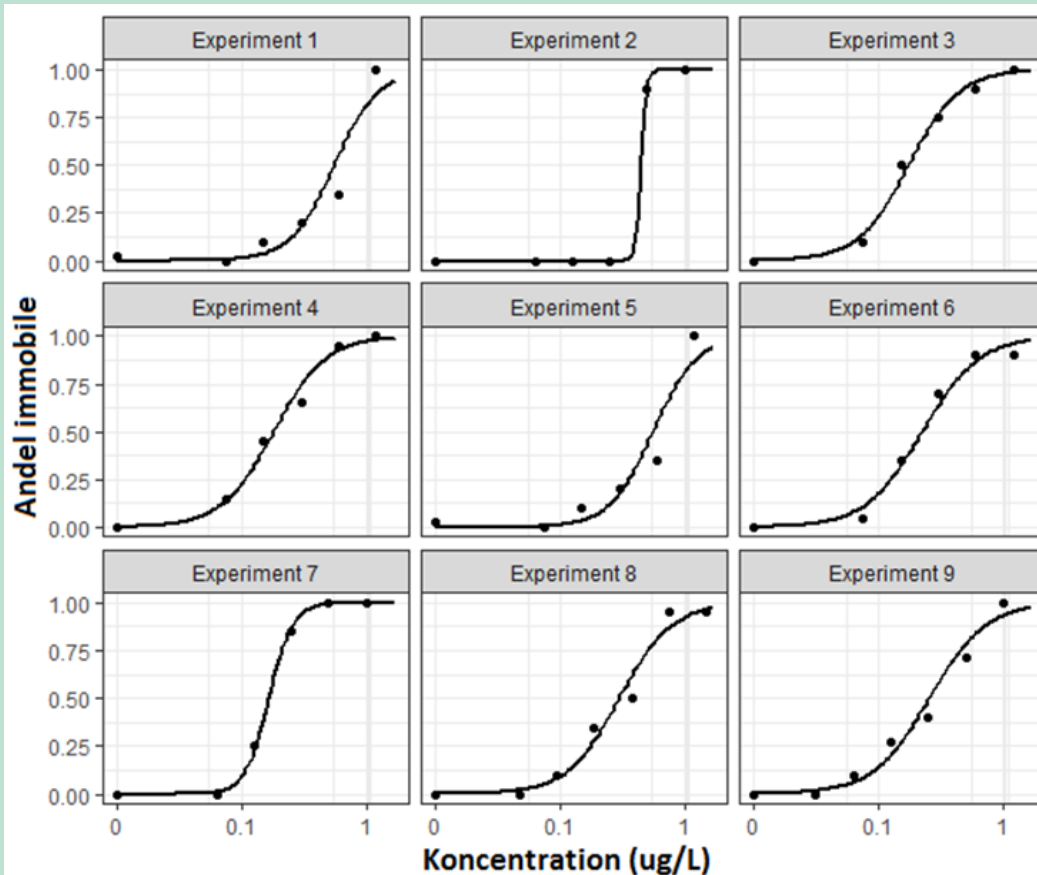
For dette eksempel var formålet at estimere BMD og BMDL for en BMR på 0,05 med "excess risk" definitionen. En to-trins meta-analyse metode blev anvendt for at tage højde for den hierarkiske struktur i data. I det første trin blev en to-parameter log-logistisk model fittet til data fra hvert deleksperiment. Fra hver model blev den estimerede BMD og den tilhørende standardfejl ekstraheret. I det andet trin blev en tilfældig-effekt meta-analytisk model estimeret til de ekstraherede BMD værdier under anvendelse af R-pakken *metafor* (Viechtbauer 2010).

Figur 8 viser de fundne modeller fra trin 1, mens tabel 4 viser de individuelt estimerede BMD og BMDL værdier.

**Tabel 4.** Estimeret Benchmark dose (BMD) med standard fejl (SE) og benchmark dose nedre grænse (BMDL) for hvert af de ni uafhængige eksperimenter med  $\alpha$ -cypermethrin.

	BMD	SE	BMDL
Experiment 1	0,17	0,10	0,04
Experiment 2	0,39	-0,58	0
Experiment 3	0,04	0,02	0,01
Experiment 4	0,04	0,02	0,01
Experiment 5	0,17	0,10	0,04
Experiment 6	0,05	0,02	0,02
Experiment 7	0,09	0,06	0,01
Experiment 8	0,07	0,04	0,02
Experiment 9	0,06	0,03	0,02





**FIGUR 8.** Koncentrations-responskurver for hver af de ni del-eksperimenter i  $\alpha$ -cypermethrin data baseret på to-parameter log-logistiske modeller.

Den estimerede BMD og BMDL, der var resultatet af den metaanalytiske model i trin 2, var henholdsvis 0,07  $\mu\text{g/L}$  og 0,05  $\mu\text{g/L}$ . Dette afviger fra standard metoderne til at kombinere data fra flere studier; et simpelt (ikke-vægtet) gennemsnit af BMD-værdierne fra de enkelte studier var 0,12  $\mu\text{g/L}$  med en standardafvigelse på 0,04, mens mindste BMD-værdi er 0,04 og mindste BMDL værdi er 0.

## 6.5 Tid-til-event data: Immobilisering af daphia som reaktion på $\alpha$ -cypermethrin

### 6.5.1 Problemstilling

Til event-tids-analyse, også kendt som overlevelsesanalyse, er den observerede respons tid til en bestemt hændelse af interesse. Inden for økotoxikologi er "tid til første reproduktion" eller "overlevelsestid" variable der spiller en afgørende rolle for hastigheden af populations tilvæksten (Begon, Townsend, and Harper 2005), og disse variable er nøglen til at forstå under hvilke betingelser denne påvirkes negativt. Inden for toksikologi kan variable såsom overlevelsestid, tid til første tumor eller tid til andre uønskede bivirkninger bruges til risikovurdering af kemikalier (OECD 2014).

Når event-tids-data skal analyseres kan en standard regressionsmodel normalt ikke bruges, da data på event-tider ofte er ufuldstændige i den forstand, at vi ikke registrerer det nøjagtige tidspunkt, hvor begivenheden af interesse (f.eks. død) finder sted. Af praktiske grunde registreres begivenheder sjældent kontinuert, men snarere på specifikke tidspunkter. Det giver

alene information om hvorvidt en bestemt begivenhed skete inden for et specifikt tidsinterval og der er således tale om intervalcensurering. Det ekstreme tilfælde af intervalcensurering er hvor det kun er muligt at undersøge, om hændelsen er sket, når organismen ofres. Hvis en hændelse er sket, ved vi så kun at den skete før observations tidspunktet, hvilket gør det til en såkaldt venstre-censureret hændelse. Hvis hændelsen ikke er sket kaldes den for højre-censureret. I sidstnævnte tilfælde ved man så kun at hændelsen sandsynligvis vil ske efter afslutningen af undersøgelsen.

I modsætning til at stole på data fra et enkelt tidspunkt, normalt status ved afslutningen af eksperimentet, gør event-tids-analyse bedre brug af de tilgængelige data, da alle observationer over tid bruges i modelleringsproceduren (Crane et al. 2016). Hvis man i stedet ignorerer informationen om hvad der skete før slutpunktet reduceres den statistiske styrke (Newman and Dixon 1996).

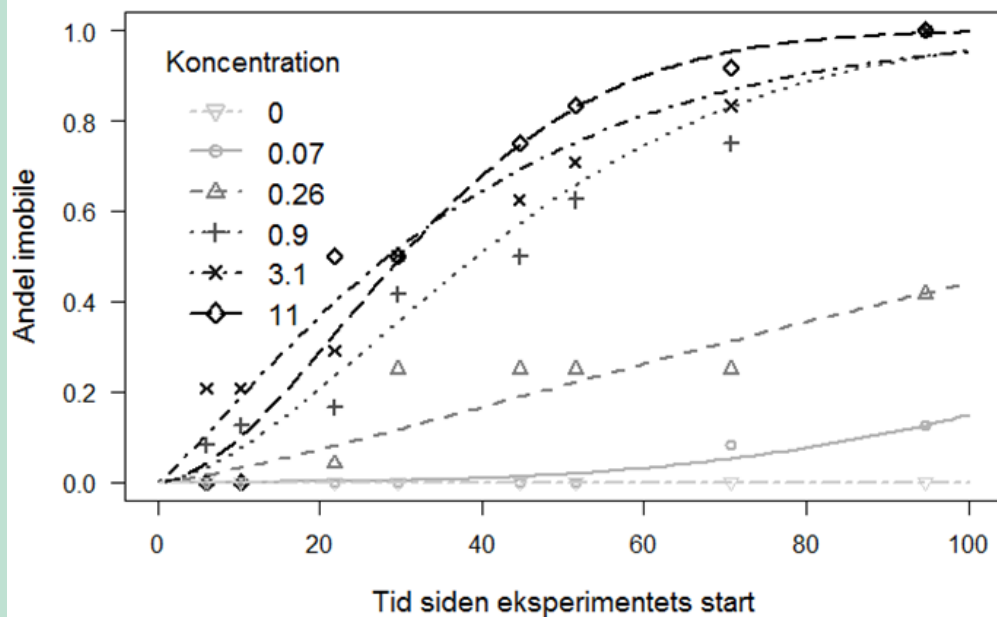
### 6.5.2 Tid til immobilisering af daphnia

Vi betragter data fra et eksperiment, der undersøgte toksikodynamikken i insekticidet  $\alpha$ -cypermethrin ved tilstedeværelsen af propiconazol i *Daphnia magna* (Cedergreen et al. 2017). Fireogtyve daphnia blev udsat for hver af seks forskellige pulskoncentrationer af  $\alpha$ -cypermethrin under tilstedeværelsen af 1,4  $\mu$ M propiconazol. De indledende  $\alpha$ -cypermethrin-koncentrationer var 0, 0,10, 0,34, 1,20, 4,08 og 14,4 nM. De gennemsnitlige faktiske pulskoncentrationer var 0, 0,07, 0,26, 0,9, 3,1 11 nM. Daphniderne blev fulgt i 94,7 timer og blev kontrolleret for immobilitet på otte forskellige tidspunkter. Disse data indeholder derfor både interval- og højrcensurerede data da ikke alle daphnia blev immobile inden for eksperimentet.

### 6.5.3 Analyse af data

Bestemmelse af BMD for event-tids-data kan foretages i to-trin. Mere specifikt, i det første trin, beskrives sandsynligheden for en hændelse som en funktion af tid ved en event-tids model der inkluderer én kurve for hver dosis. I det andet trin fittes en dosis-respons-model (for kontinuerede data) til estimerne af  $t_{50}$  baseret på modellen i trin 1. Givet et foruddefineret BMR estimeres BMD derefter fra den resulterende model i trin 2.

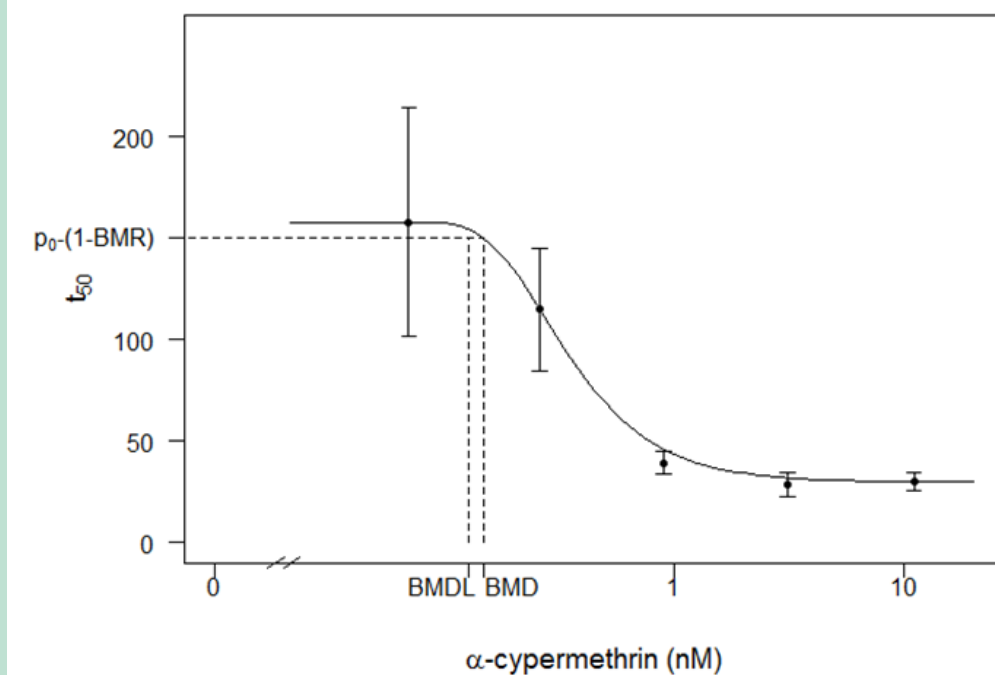
For kontroldosis, 0 nM, var alle daphnider stadig mobile efter afslutningen af eksperimentet. Det var derfor ikke muligt at estimere en dosis-respons-model for kontrollerne. For de resterende fem doser blev der estimeret en fælles to-parameter Weibull-model med en event-tids kurve for hver dosis. Den resulterende model er vist i figur 9, hvor det ses, at event-tids modellen giver en rimelig beskrivelse af data for alle fem aktive doser. For to af kurverne ville en model der tillod en øvre grænse under 1 højst sandsynligt beskrive data bedre; for 0,26 nM-koncentrationen, hvor det seneste tidspunkt ser for højt ud og for 0,07-koncentrationen hvor der er en stigning omkring 60 nM med et plateau efter. Uden at vide hvad der skete efter det seneste tidspunkt kunne vi imidlertid ikke blot kassere nogen af datapunkterne som fejl. Flere observationer i slutningen af eksperimentet ville muligvis have gjort det muligt at beskrive forløbet for i især disse to koncentrationer bedre og ville højst sandsynligt have muliggjort brug af en tre-parameter model med inklusion af et øvre plateau. En længere opfølgingsperiode ville nok have vist at mortaliteten ville flade ud, da erfaring viser at den toksikologiske effekt fra puls forsvinder efter ca. 5 dage hvorefter baggrundsmortaliteten tager over. Som data var her gav en tre-parameter model ikke brugbare estimer for øvre grænse ( $>1$ ), hvorfor to-parametermodellen blev bibeholdt.



**FIGUR 9.** Weibull event-tids modellen fittet til daphnia-immobiliseringsdata for alle fem aktive doser af  $\alpha$ -cypermethrin (trin 1 i to trins metoden). Da alle daphnia i kontrolgruppen forblev mobile i løbet af den eksperimentelle periode, blev der ikke estimeret nogen model for kontrolterne. I stedet er der tilføjet en flad grå to-stiplet linje for illustration.

For hver dosis blev  $t_{50}$  og den tilhørende kombinerede kovariansmatrix udledt fra den fittede model. Da  $t_{50}$  er en afledt parameter for Weibull-modellen, dvs. en parameter der ikke indgår direkte i modelbeskrivelsen, blev den multivariate delta-metode anvendt til at estimere den afledte parameter vektor og den tilknyttede kovariansmatrix.

I det andet trin blev en fire-parameter Weibull-model hvor både øvre og nedre grænse er estimeret brugt til at beskrive sammenhængen mellem  $t_{50}$ -estimerne og dosis. Kovariansmatricen indgik som vægt i estimeringen af modellen således at  $t_{50}$ -estimer med stor usikkerhed fik mindre indflydelse på den endelige model. Den resulterende model er vist i figur 10.



**FIGUR 10.** Weibull dosis-respons-model fittet i trin 2 til  $t_{50}$  for de aktive doser af  $\alpha$ -cypermethrin. Placeringen af den estimerede BMD og BMDL for et BMR = 0,05 under anvendelse af den relative definition er indikeret med lodrette stiplede linjer. Punkter angiver de estimerede  $t_{50} \pm$  SE fundet i trin 1.

Da der i trin 2 ikke var tilgængelige data for kontrol dosis kunne hybridmetoden ikke anvendes til at finde BMD. I stedet blev BMD bestemt med den relative definition og BMD og BMDL fundet for BMR = 0,05. Det betyder, at BMD var den dosis, der var forbundet med en 5% ændring i  $t_{50}$  i forhold til den estimerede øvre grænse. Den estimerede BMD og BMDL var henholdsvis 0,15 og 0,13 nM  $\alpha$ -cypermethrin.

Havde vi i stedet ignoreret tidsforløbet gennem forsøget og kun brugt andelen af daphnia som var immobile ved afslutningen af forsøget, kunne data analyseres ved hjælp af en to-parameter Weibull-model under antagelse af binomiale responsdata. På baggrund af denne model, med BMR = 0,05 og excess risk definitionen blev BMD og BMDL henholdsvis 0,05 og 0,002 nM  $\alpha$ -cypermethrin. I dette tilfælde, hvor nedre og øvre grænse er henholdsvis 0 og 1, kan excess risk definition direkte oversættes til den relative ændring i respons der er brugt i to trins modellen. Det vil sige, sammenlignet med analysen, der anvender alle data, resulterer den naive tilgang der kun bruger data fra udgangen af studiet i en lavere BMD og en BMDL længere væk fra den estimerede BMD.

## 6.6 SSD

### 6.6.1 Problemstilling

Arts følsomhedsfordelinger, eller Species Sensitivity Distributions (SSD'er), anvendes til at fastsætte grænseværdier for økosystemer. De bliver som oftest anvendt til fastsættelse af grænseværdier for pesticider i akvatiske miljøer, hvor de har vist sig at beskrive økosystem-responsen lige så godt som de mere omfattende mikrokosmos eksperimenter, der undersøger effekter på mini-økosystemer (Posthuma, Suter II, and Traas 2002). SSD'er viser fordelingen

af et udvalg af arters følsomhed (minimum 8 arter) over for et bestemt kemikalie normalt målt som en EC-værdi eller en NOAEC. Fordelingen antages at være repræsentativ for organismerne i økosystemet og at være log-normal fordelt, og beskrives derfor ofte med en log-normal eller log-logistisk model. Man beregner så den koncentration hvor 95% af populationen ikke er påvirket (til det niveau, som man bruger som mål for følsomhed). Det kaldes en HC5 (5% Hazard Concentration) og HC5 eller dens nedre konfidensgrænse (95%) anvendes som udgangspunkt, når der skal udarbejdes grænseværdier for det pågældende stof.

Når der skal laves SSD anvendes som beskrevet EC eller NOAEC-estimer, men usikkerheden med hvilken de er bestemt tages traditionelt set ikke med i betragtning når SSD fordelingen skal findes. En måde hvorpå man kan imødekomme dette er at bruge BMDL-estimer som udgangspunkt for bestemmelsen af SSD funktionen. På den måde tages der højde for en del af usikkerheden relateret til bestemmelsen af effekt-estimatet (som i dette tilfælde vil være BMD).

### 6.6.2 Data

Data til dette eksempel stammer fra en artikel der undersøger den akvatiske toksicitet af forskellige saponin-rige ekstrakter brugt som biopesticider (Xiaogang Jiang et al. 2018). Vi betragter tre forskellige saponin ekstrakter; saponin ekstrakt fra quillaja bark, te frøskaller og quinoa frøskaller, samt et rent quillaja saponin. Tolv forskellige vandlevende organismer blev brugt som test organismer for de fire saponiner i akutte toksicitets test. Antallet af koncentrationer varierede fra 5 til 9 alt efter test-organisme og saponin. For to af organismerne (*Raphidocelis subcapitata* og *Lemna minor*) var det undersøgte respons relativ vækstrate. For tre organismer (*Daphnia magna*, *Chironomus riparius* og *Chaoborus crystallinus*) var respons immobilisering. For *Aliivibrio fischeri* blev der undersøgt hæmning af luminescens, mens de resterende seks organismer (*Gammarus pulex*, *Tubifex tubifex*, *Lumbriculus variegatus*, *Lymnaea stagnalis*, *Lymnaea stagnalis embryo* samt *Danio rerio embryo*) blev observeret for mortalitet. Respons er således en blanding af binomiale og kontinuerte data. Ikke alle kombinationer af saponin og organisme var tilstede i data.

### 6.6.3 Analyse af data

For hver organisme og saponin kombination tilgængelig i data blev der estimeret en dosis-responskurve. For hæmning af luminescens blev der antaget en to-parameter log-logistisk model, for relative vækst rater en tre-parameter log-logistisk model, mens mortalitet og immobilitet blev modelleret med en to-parameter log-logistisk model med antagelsen om binomiale respons data.

For hver model, blev LC<sub>50</sub>/EC<sub>50</sub> og BMD svarende til BMR = 10% bestemt. For de binomiale responsdata blev BMD bestemt baseret på "additional risk" definitionen. For de kontinuerte respons data blev BMD baseret på hybrid metoden med baggrundssandsynligheden defineret ud fra en ændring på 2 standard deviationer for kontrolgruppen. BMDL blev fundet ved ikke-parametrisk bootstrap for modeller med kontinuerte responsdata og parameterisk bootstrap for modeller med binomiale respons data. Tabel 5 og 6 viser henholdsvis de resulterende LC<sub>50</sub>/EC<sub>50</sub> samt BMD og BMDL estimer.

Baseret på de fundne EC<sub>50</sub>/LC<sub>50</sub> og BMDL blev der nu lavet en SSD analyse. Vi brugte en generaliseret log-logistisk model (en fem-parameter Burr type III, med to parametre holdt fast) til at beskrive den kumulerede fordeling af henholdsvis EC<sub>50</sub> og BMDL. LC<sub>50</sub>/EC<sub>50</sub>-værdier uden for dosis-området blev ikke taget i betragtning i analyserne, men indgik i rangeringen. I praksis betød det at LC<sub>50</sub>/EC<sub>50</sub>-værdier >1000 fik tildelt værdien 1000, som et konservativt estimat, til beregning af fordelingsfunktionen, men derefter blev de ignoreret. Figur 11-14 viser de estimerede fordelingsfunktioner.

De resulterende fordelingsfunktioner beskriver både fordelingen af LC<sub>50</sub>/EC<sub>50</sub>-værdier og BMDL værdier ganske fint. Det skal bemærkes at det var muligt at basere SSD analyserne på flere organismer hvis BMDL blev brugt frem for LC<sub>50</sub>/EC<sub>50</sub>. Det skyldes i særdeleshed at

førstnævnte kun kræver en god bestemmelse af den nedre del af dosis-respons kurven, mens et  $EC_{50}$  estimat kun bestemmes præcist hvis også øvre grænse er godt bestemt.

**TABEL 5.** Estimeret EC<sub>50</sub>/LC<sub>50</sub> med tilhørende 95%-konfidensinterval (CI) for tolv organismer og fire forskellige saponiner

	Quillaja saponins		Tea saponins		Quinoa saponins		Quil-A	
	EC <sub>50</sub>	(95% CI)	EC <sub>50</sub>	(95% CI)	EC <sub>50</sub>	(95% CI)	EC <sub>50</sub>	(95% CI)
<i>Aliivibrio fischeri</i>	942,8*	(618,5 - 1267,0)	-		-		524,1*	(406,8 - 641,4)
<i>Raphidocelis subcapitata</i>	216,7	(153,9 - 279,6)	121,1	(110,1 - 132,1)	-		807,3*	(619,9 - 994,6)
<i>Lemna minor</i>	152,9	(119,6 - 186,3)	106,1	(91,3 - 120,9)	880,9*	(144,0 - 1617,9)	92,8	(65,5 - 120,1)
<i>Daphnia magna</i>	30,5	(23,7 - 37,4)	26,4	(17,9 - 35,0)	280,9	(-1684,3 - 2246,1)	19,3	(16,3 - 22,4)
<i>Chironomus riparius</i>	465,6	(233,6 - 697,5)	180,7	(121,7 - 239,7)	416,1	(342,9 - 489,2)	107,6	(91,6 - 123,6)
<i>Chaoborus crystallinus</i>	44024,8*	(-173098,0 - 261147,6)	-		-		26379,1*	(-95256,2 - 148014,3)
<i>Gammarus pulex</i>	639,3*	(-1208,9 - 2487,4)	258,5*	(-15,2 - 532,3)	-		114,1	(52,0 - 176,2)
<i>Tubifex tubifex</i>	70,8	(59,9 - 81,7)	4,8	(4,0 - 5,7)	186,5	(153,1 - 219,8)	37,3	(33,0 - 41,6)
<i>Lumbriculus variegatus</i>	59,6	(47,9 - 71,3)	3,4	(3,0 - 3,8)	126,3	(101,4 - 151,2)	39,0	(-609,1 - 687,1)
<i>Lymnaea stagnalis</i>	78,2	(-1220,3 - 1376,6)	3,8	(2,0 - 5,7)	97,3	(-1522,6 - 1717,2)	32,5	(-19,3 - 84,2)
<i>Lymnaea stagnalis embryo</i>	55,0	(44,1 - 66,0)	18,6	(14,0 - 23,3)	72,9	(52,3 - 93,4)	18,7	(13,2 - 24,2)
<i>Danio rerio embryo</i>	13,4	(9,6 - 17,2)	3,4	(2,4 - 4,3)	131,9	(92,7 - 171,2)	11,9	(8,3 - 15,4)

- Der foreligger ingen data

\* Den estimerede koncentration ligger udenfor dosis området

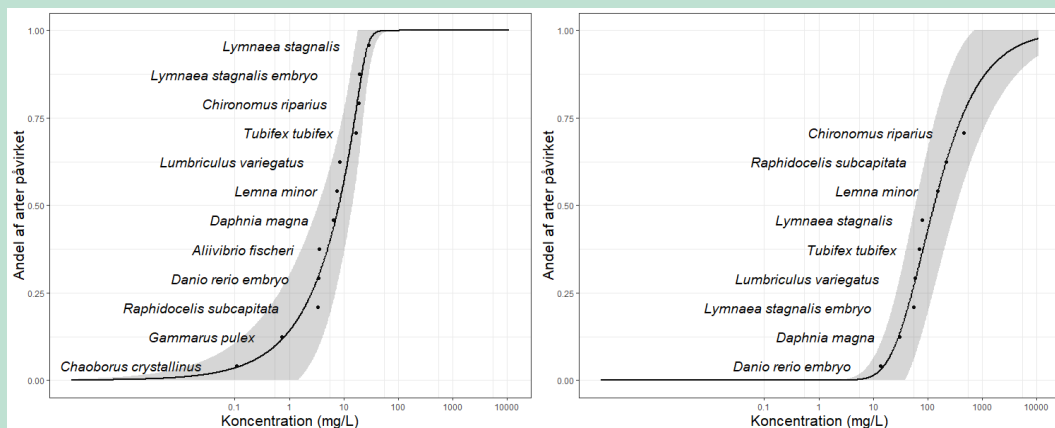
**TABEL 6:** Estimeret BMD<sub>10</sub> og BMDL<sub>10</sub> associeret med en additiv ændring i respons (BMR) på 10% for tolv organismer og fire forskellige saponiner

	Quillaja saponins		Tea saponins		Quinoa saponins		Quil-A	
	BMD10	BMDL10	BMD10	BMDL10	BMD10	BMDL10	BMD10	BMDL10
<i>Aliivibrio fischeri</i>	12,6	3,5	-	-	-	-	21,7	11,6
<i>Raphidocelis subcapitata</i>	5,0	3,4	36,1	22,6	-	-	284,1	219,7
<i>Lemna minor</i>	13,5	7,6	21,4	14,6	8,7	3,7	6,2	3,0
<i>Daphnia magna</i>	12,7	6,5	5,3	2,2	272,6	140,4	14,0	6,0
<i>Chironomus riparius</i>	77,7	18,8	50,9	12,6	261,0	91,3	56,9	34,1
<i>Chaoborus crystallinus</i>	12,8	0,1	-	-	-	-	14,7	0,8
<i>Gammarus pulex</i>	12,0	0,7	31,0	7,4	-	-	21,1	9,4
<i>Tubifex tubifex</i>	39,8	16,7	3,9	0,8	97,9	42,9	24,2	10,4
<i>Lumbriculus variegatus</i>	37,7	8,5	2,9	0,006	67,2	33,8	37,2	15,1
<i>Lymnaea stagnalis</i>	74,5	28,8	3,7	1,6	92,7	36,1	29,3	14,0
<i>Lymnaea stagnalis embryo</i>	32,2	19,4	8,3	5,5	30,2	14,2	6,1	3,4
<i>Danio rerio embryo</i>	7,0	3,4	2,1	0,9	63,9	26,9	5,7	2,9

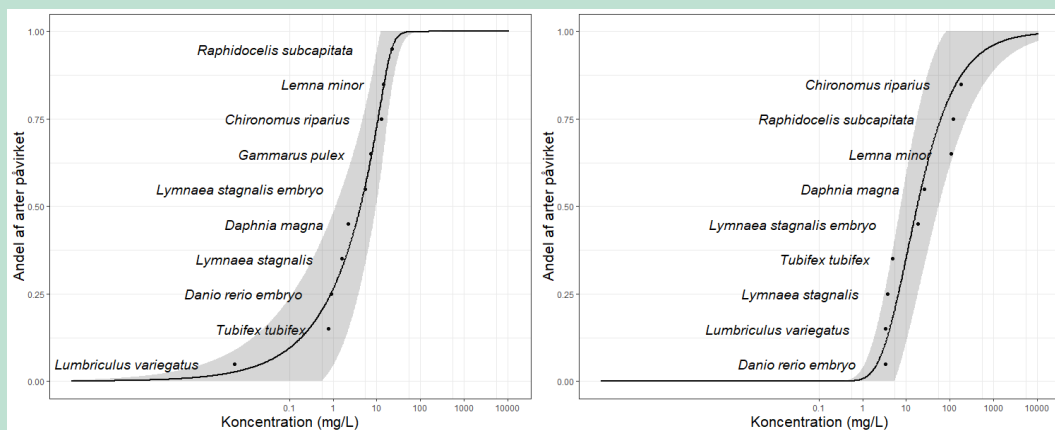
- Der foreligger ingen data

\* Den estimerede koncentration ligger udenfor dosis området

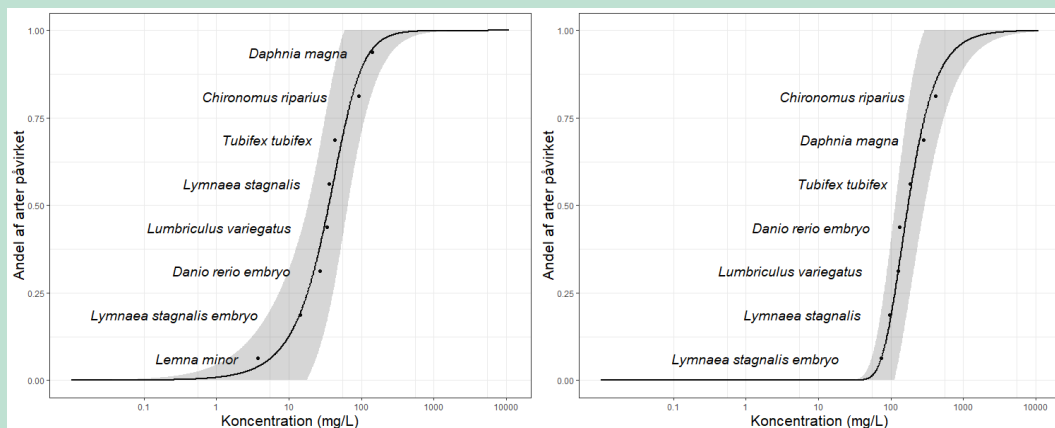




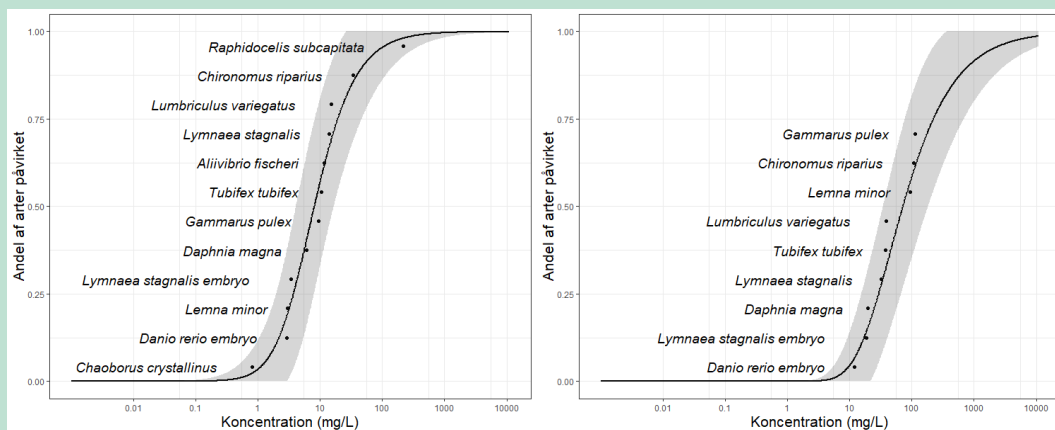
**FIGUR 11.** Estimerede kurver for art følsomheds fordelingen for quillaja saponiner baseret på BMDL<sub>10</sub> (til venstre) og EC<sub>50</sub>/LC<sub>50</sub> (til højre).



**FIGUR 12.** Estimerede kurver for art følsomheds fordelingen for te saponiner baseret på BMDL<sub>10</sub> (til venstre) og EC<sub>50</sub>/LC<sub>50</sub> (til højre).



**FIGUR 13.** Estimerede kurver for art følsomheds fordelingen for quinoa saponiner baseret på BMDL<sub>10</sub> (til venstre) og EC<sub>50</sub>/LC<sub>50</sub> (til højre).



**FIGUR 14.** Estimerede kurver for art følsomheds fordelingen for Quil-A baseret på BMDL<sub>10</sub> (til venstre) og EC<sub>50</sub>/LC<sub>50</sub> (til højre).

For begge SSD for alle fire saponiner blev hazard koncentrationerne (HC) defineret som koncentrationerne der resulterer i at henholdsvis 5, 10 og 50% af organismerne er påvirket estimeret ud fra de modellerede fordelingsfunktioner. Resultaterne ses i tabel 7. Som forventet ligger HC-værdierne baseret på BMDL meget lavere end HC-værdierne baseret på EC<sub>50</sub>/LC<sub>50</sub>. Det skyldes både at BMDL er fundet med et BMR på 10% i forhold til de 50% brugt for EC<sub>50</sub>/LC<sub>50</sub>, og at BMDL inkorporerer usikkerheden i estimatet. Vi bemærker, at forskellene mellem HC-værdierne fundet på baggrund af de to typer af DOD varierer fra en faktor ca. 5 til en faktor ca. 100.

**TABEL 7:** Estimerede hazard koncentrationer for 5 (HC<sub>5</sub>), 10 (HC<sub>10</sub>) og 50 (HC<sub>50</sub>) procent organismer påvirket, beregnet ud fra arts følsomheds fordelingskurver baseret på henholdsvis ED<sub>50</sub> eller BMDL<sub>10</sub>.

		Quillaja saponiner	Te saponiner	Quinoa saponiner	Quil-A
HC <sub>5</sub>	ED <sub>50</sub>	17,2 (8,9 – 52,7)	2,1 (1,1 – 7,3)	70,0 (51,6 – 125,6)	10,4 (5,8 – 28,7)
	BMDL <sub>10</sub>	0,2 (0,02 – 2,5)	0,02 (0,001– 1,05)	4,6 (1,1 – 22,5)	1,2 (0,5 – 4,0)
HC <sub>10</sub>	ED <sub>50</sub>	25,2 (13,2 – 67,5)	3,2 (1,6 – 9,4)	82,7 (60,2 – 139,4)	14,8 (8,1 – 36,3)
	BMDL <sub>10</sub>	0,6 (0,09 – 3,6)	0,1 (0,008 – 1,6)	8,2 (2,4 – 27,3)	1,9 (0,9 – 4,9)
HC <sub>50</sub>	ED <sub>50</sub>	131,1 (59,1 – 337,8)	17,8 (7,4 – 55,2)	172,6 (114,5 – 291,7)	71,6 (34,2– 181,3)
	BMDL <sub>10</sub>	8,1 (3,4 – 14,3)	4,2 (1,1 – 9,6)	35,5 (18,3 – 64,8)	8,0 (4,1 – 16,4)

## 6.7 Opsummering

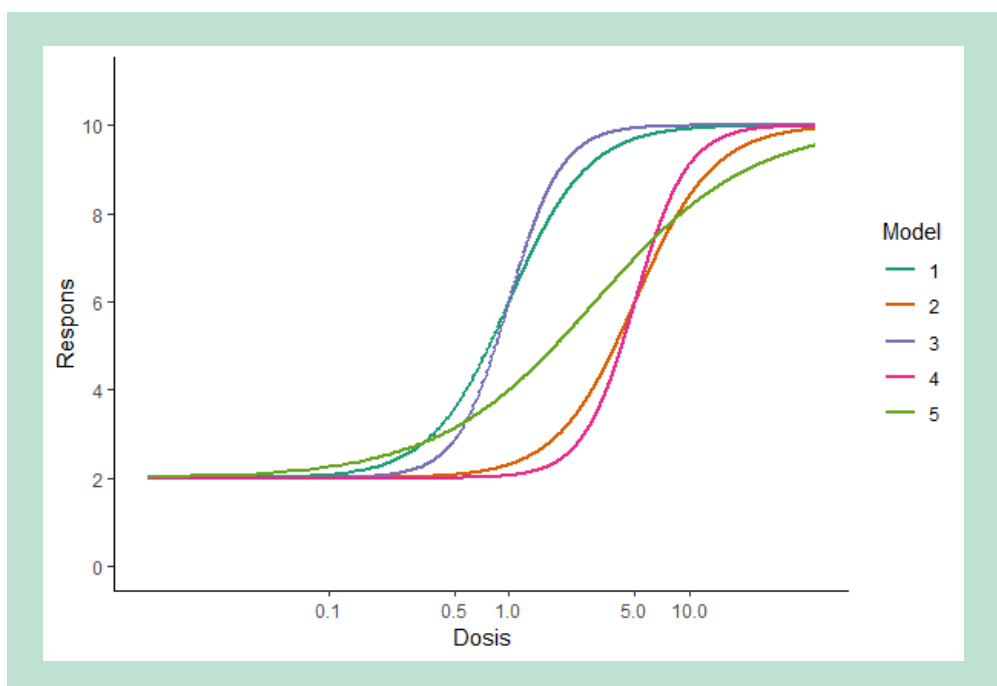
De seks eksempler gennemgået her illustrerer fleksibiliteten af BMD-metoden til at håndtere forskellige analytiske problemstillinger relateret til data fra økotoxikologien. Foruden de to mest gængse data typer, for binomiale og kontinuerte data, blev det desuden vist hvordan man kan finde BMD og BMDL for tælle data og tid-til-event data. De to data typer optræder ofte i økotoxikologien, men bliver sjældent analyseret optimalt. Endelig blev det illustreret at SSD-analyser fra et statistisk synspunkt kan baseres på BMDL som for EC<sub>50</sub> (eller EC<sub>x</sub>, NOAEL eller LOAEL), hvilket vil give mulighed for at tage højde for usikkerheden i data i højere grad end tidligere.

# 7. Sammenligning af forskellige metoder til bestemmelse af BMDL

Der er flere metoder til at bestemme BMDL. I det følgende afsnit sammenlignes metoderne, som er tilgængelige i *bmd*-pakken baseret på simulationer.

## 7.1 Opsætning af simulationer

For at sammenligne de forskellige metoder blev der genereret data fra fem forskellige modeller. Vi benyttede en fire-parameter log-logistisk model med fem forskellige sæt af parametre. I alle modeller var nedre og øvre grænse antaget at være henholdsvis 2 og 10. Hældningsparameteren og  $ED_{50}$  var i model 1 til 5 som følger: (-2,1), (-2,5), (-3,1), (-3,5) og (-1, 3). Det dækker over to sæt af kurver med samme hældning placeret forskellige steder og en ekstra kurve med en noget fladere hældning. De fem modeller er illustreret i Figur 15.

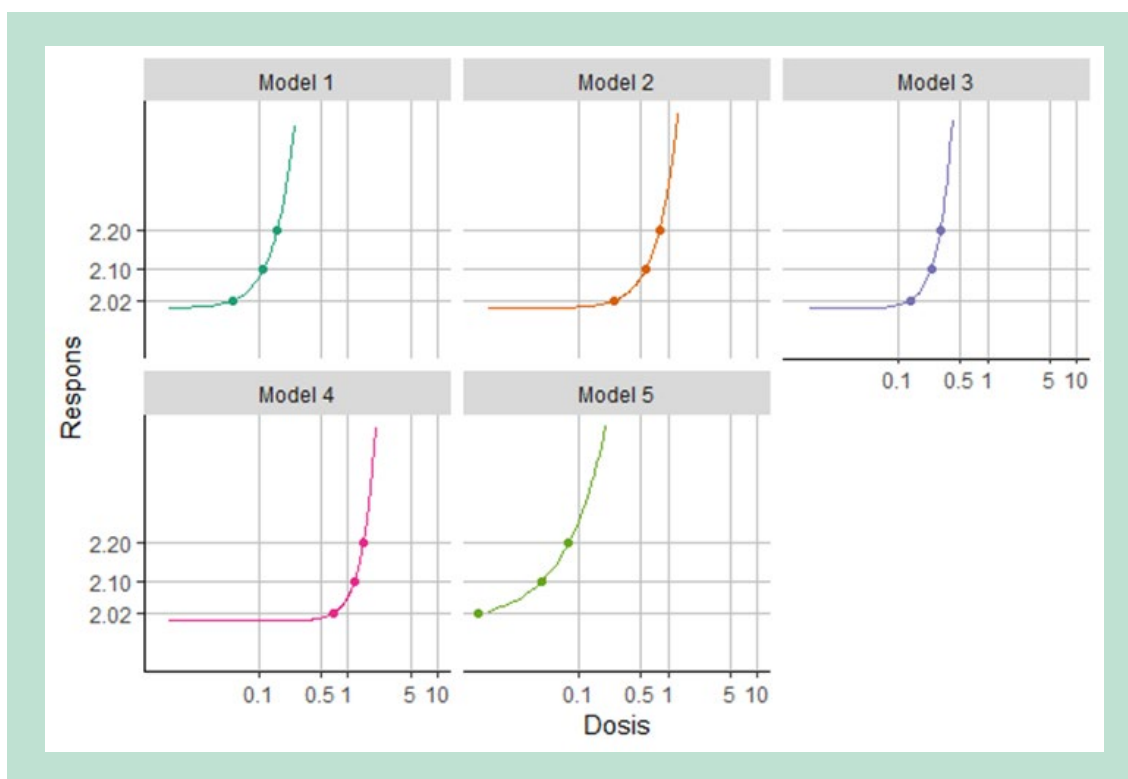


**FIGUR 15.** De fem log-logistiske modeller der bruges som udgangspunkt for simulationerne. Markeringerne på x-aksen indikerer dosis-placeringerne.

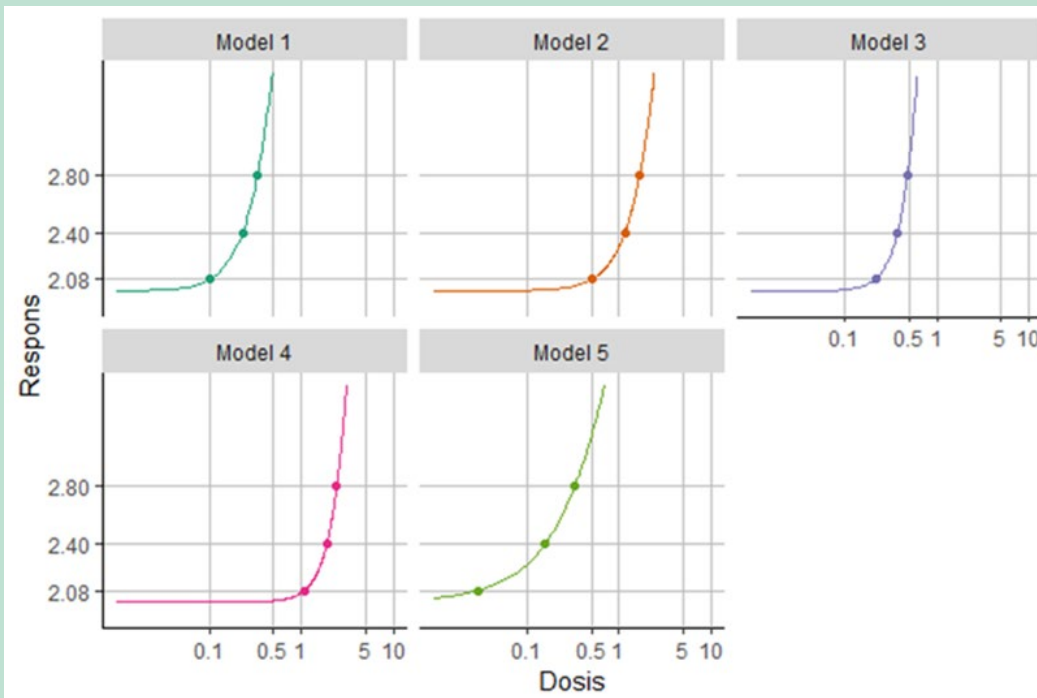
Data blev genereret under en antagelse om normalfordelte data med en standard afvigelse på enten 0,1 eller 1. Der blev genereret data fra seks forskellige doser (0, 0,1, 0,5, 1, 5, 10), med enten 10 eller 100 gentagelser for hver dosis.

For hver kombination af model, standard afvigelse og antal gentagelser blev der genereret 1000 datasæt. For hvert af disse datasæt blev BMD og BMDL svarende til BMR = 0,01, 0,05 og 0,1 bestemt for henholdsvis den "relative" og "extra" definitionen, mens der for "added" definitionen blev benyttet BMR = 0,01, 0,1 og 1,0 for at ramme nogenlunde de samme absolutte ændringer som for "relative" og "extra". For de betragtede modeller svarer BMR = 0,01, 0,05

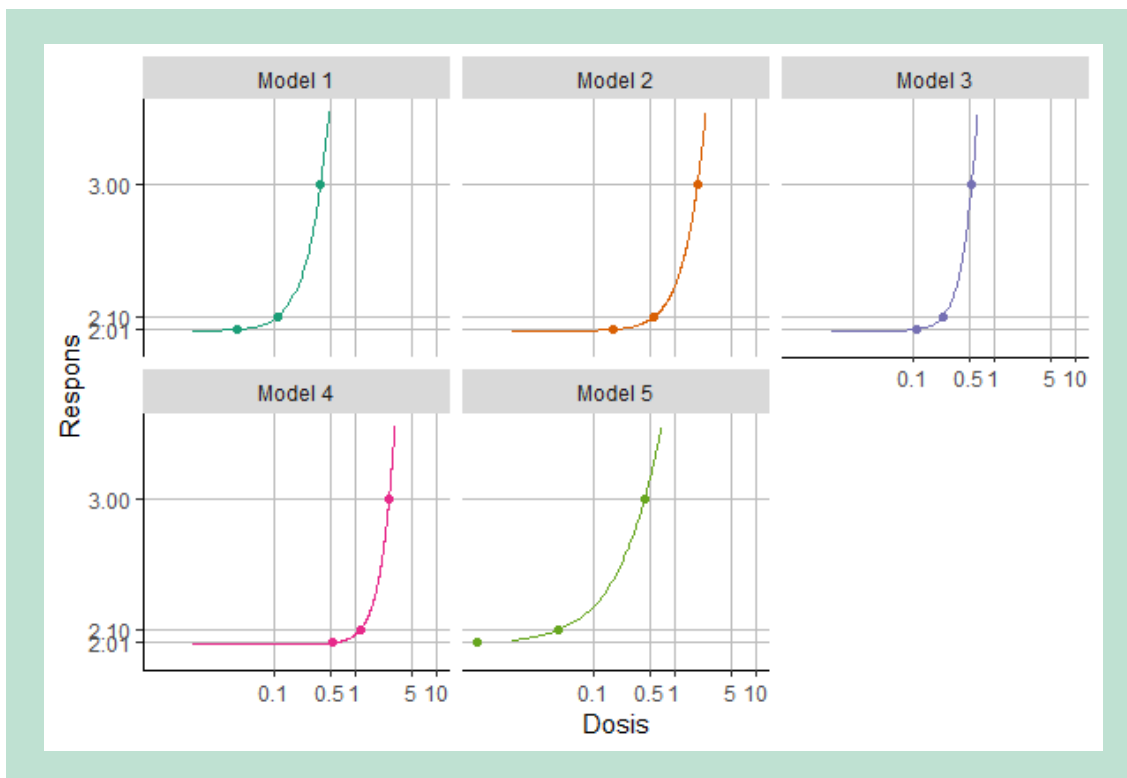
og 0,1 for "relative" og "extra" definitionerne til absolutte ændringer på henholdsvis 0,02, 0,10 og 0,20 samt 0,08, 0,50 og 0,80. Placering af BMD for hver model og hver værdi af BMR kan ses på figur 16 for den relative definition, 17 for "extra" definitionen og 18 for "added" definitionen. Man kan se, at nogle af de anvendte modeller resulterer i BMD-værdier i dosis-områder tæt på data mens andre rammer i dosis-områder hvor der ikke er mange data.



**FIGUR 16.** Placering af de tre BMD-værdier hørende til de tre BMR-værdier 0,01, 0,05 og 0,1 ved brug af den relative definition for hver af de fem modeller brugt i simulationerne. Lodrette grå streger indikerer dosis-placeringerne, mens de vandrette grå streger indikerer respons niveauet der svarer til en relative ændring fra baggrunds niveauet på henholdsvis 1%, 5% og 10%.



**FIGUR 17.** Placering af de tre BMD-værdier hørende til de tre BMR-værdier 0,01, 0,05 og 0,1 ved brug af "extra" definitionen for hver af de fem modeller brugt i simulationerne. Lodrette grå streger indikerer dosis-placeringerne, mens de vandrette grå streger indikerer responsniveauet der svarer til en ændring fra baggrundsniveauet baseret på extra definitionen på henholdsvis 1%, 5% og 10%



**FIGUR 18.** Placering af de tre BMD-værdier hørende til de tre BMR-værdier 0,01, 0,10 og 1,0 ved brug af "added" definitionen for hver af de fem modeller brugt i simulationerne. Lodrette grå streger indikerer dosis-placeringerne, mens de vandrette grå streger indikerer respons niveauet der svarer til en absolut ændring fra baggrundsniveauet på henholdsvis 0,01, 0,10 og 1,0.

For at undersøge konsekvensen af at antagelsen om varianshomogenitet ikke var opfyldt blev der desuden for "extra" definitionen genereret data, hvor der for dosisniveauerne 1 og 5 blev tilføjet yderligere støj i form af normalfordelte værdier med standard afvigelse på henholdsvis samme og halvdelen af den standard afvigelse der var brugt til at generere data.

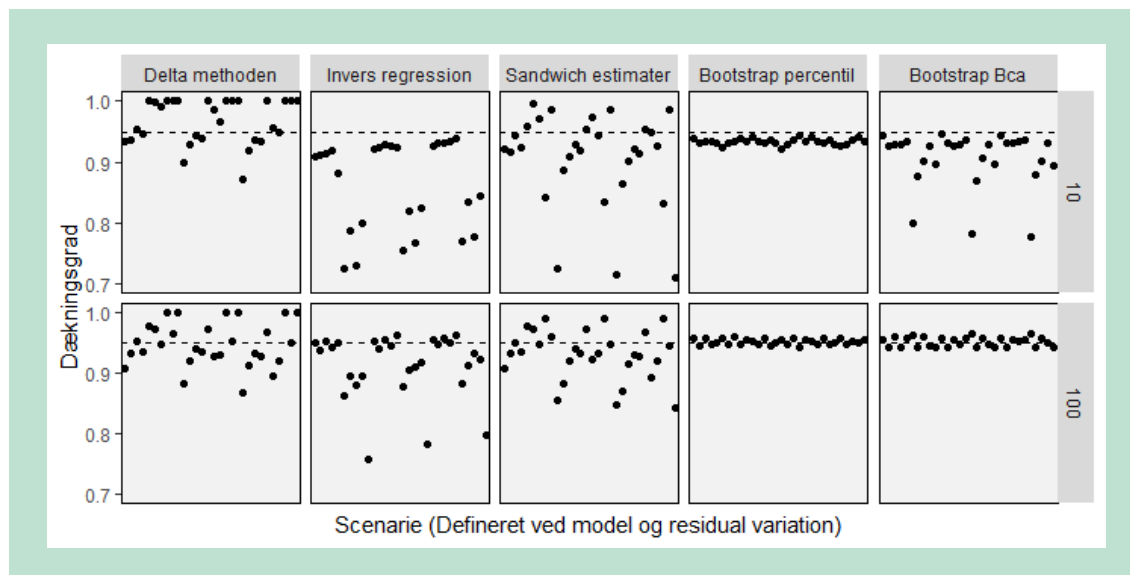
I det følgende opsummeres resultaterne af simulationerne under estimation af den "sande" model. Det vil sige, samme type model data er genereret fra (log-logistisk) blev også brugt til at estimere dosis-respons sammenhængen for hvert genereret datasæt. Dækningsgraden af BMDL blev bestemt som andelen af de 1000 datasæt for hvilke det sande BMD (bestemt ud fra modellen data var genereret fra) lå over BMDL. Denne dækningsgrad skulle gerne ligge på 95%.

## 7.2 Resultater

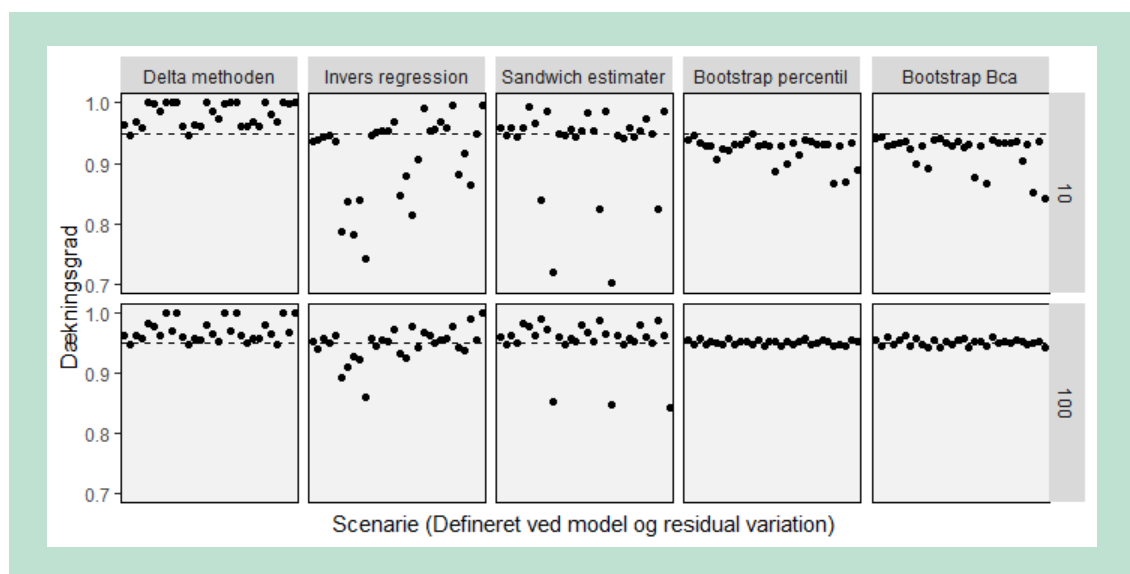
### 7.2.1 Effekten af antal gentagelser

Figur 19, 20 og 21 viser dækningsgrader for de fem metoder, opdelt på antal gentagelser for hver dosis for henholdsvis den relative, "extra" og "added" definitionen. Ikke overraskende blev resultaterne for alle fem metoder bedre ved et større antal gentagelser. Det gjaldt blandt andet for delta-metoden hvor dækningsgraderne for de små datasæt ofte lå på 100%, hvilket skyldes alt for brede konfidensintervaller omkring BMD og dermed voldsomt konservative estimater af BMDL. For begge bootstrap metoder betød det øgende antal observationer, at dækningsgraderne lagde sig tættere på det nominelle niveau. Især fik BCa metoden for den relative defini-

tion ramt rigtigt for nogle af de scenarier der før ledte til alt for lave dækningsgrader. For sandwich estimatorne blev dækningsgraderne for de store datasæt højere for de scenarier, der for de små datasæt resulterede i meget lave dækningsgrader, men de var stadig noget under det nominelle niveau. For invers blev dækningsgraden for de fleste scenarier rykket tæt på det nominelle niveau men stadig med enkelte scenarier der havde dækningsgrader langt under det nominelle niveau.

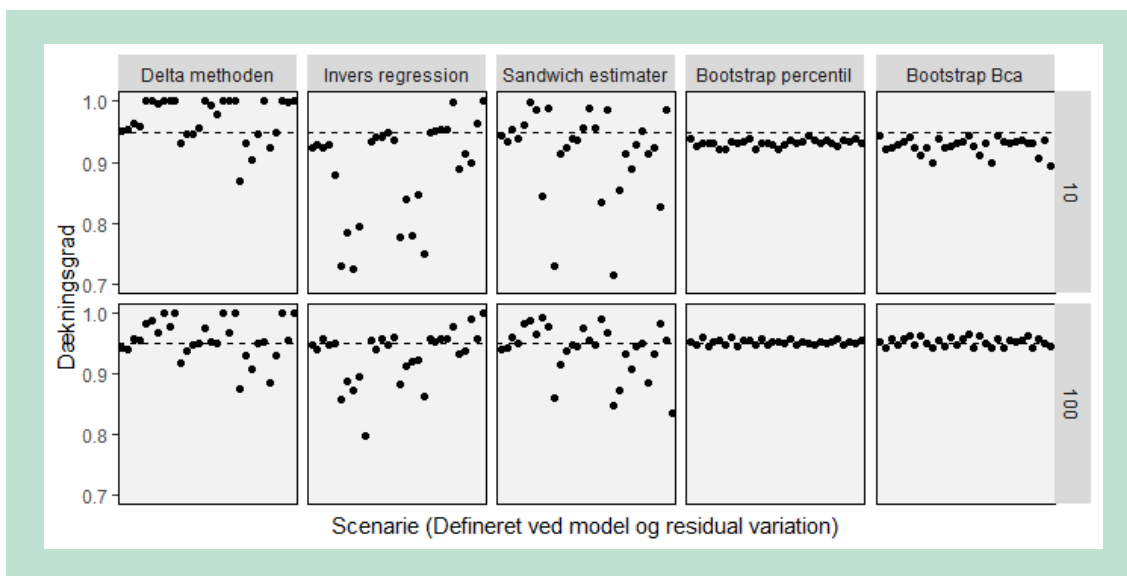


**FIGUR 19.** Dækningsgrader af BMDL for fem forskellige metoder opdelt på antal gentagelser inden for hver dosis, N=10 eller 100, når BMD er defineret ved den relative definition.



**FIGUR 20.** Dækningsgrader af BMDL for fem forskellige metoder opdelt på antal gentagelser inden for hver dosis, N=10 eller 100, når BMD er defineret ved "extra" definitionen.

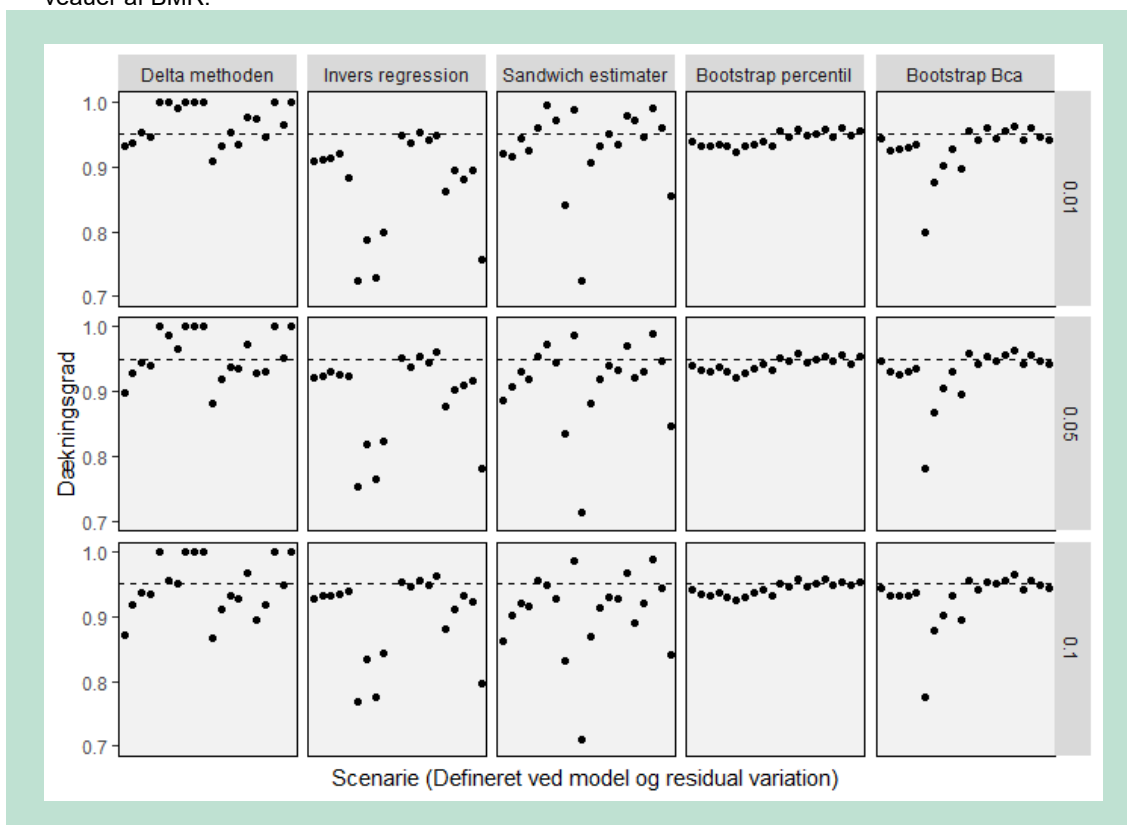




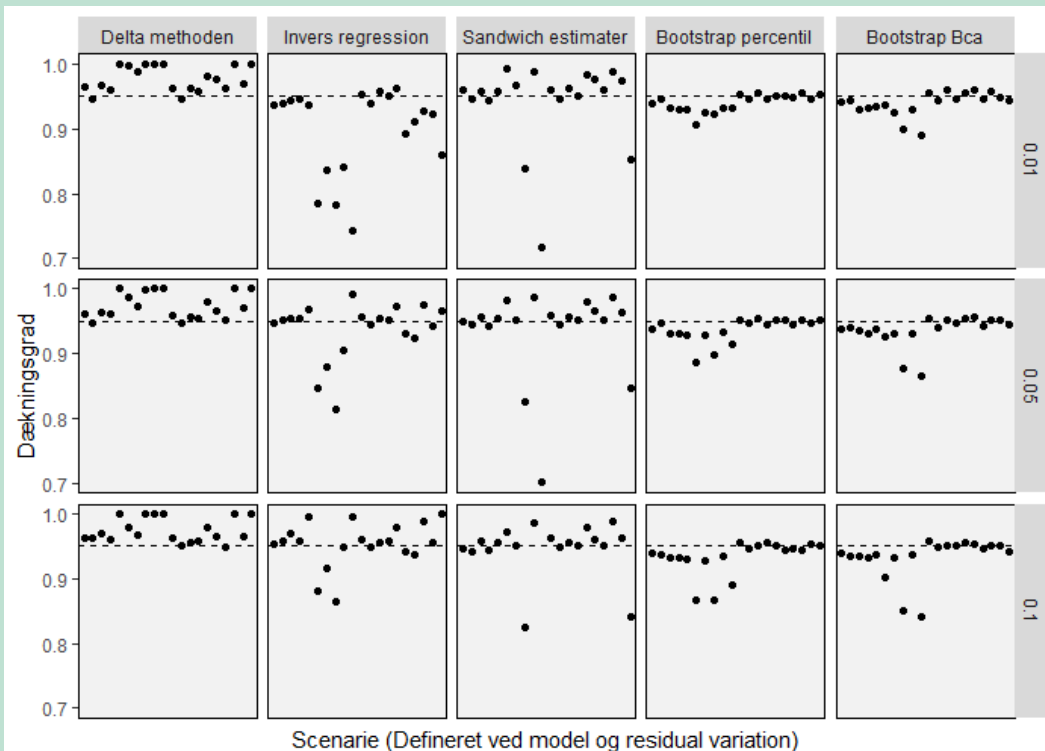
**FIGUR 21.** Dækningsgrader af BMDL for fem forskellige metoder opdelt på antal gentagelser inden for hver dosis, N=10 eller 100, når BMD er defineret ved "added" definitionen.

### 7.2.2 Størrelsen på BMR

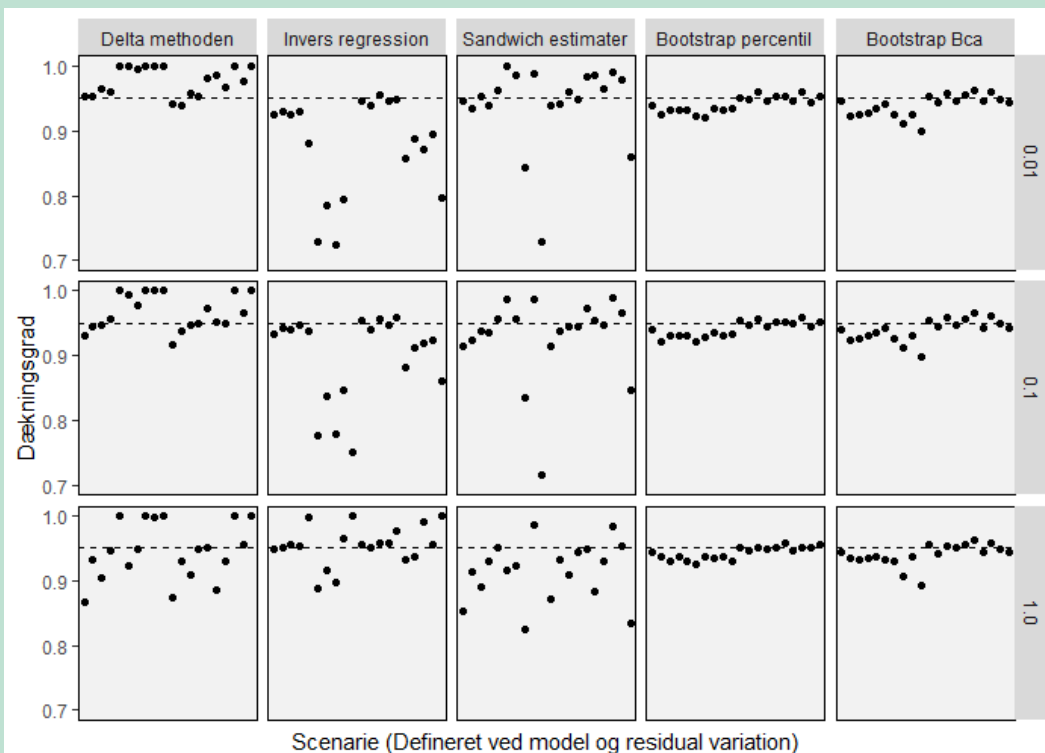
Figur 22, 23 og 24 viser dækningsgrader delt op på størrelsen af BMR for henholdsvis den relative, "extra" og "added" definitionen. For alle metoder ser vi samme mønster på tværs af niveauer af BMR.



**FIGUR 22.** Dækningsgrader af BMDL for fem forskellige metoder opdelt på størrelsen af BMR, 0,01, 0,05 og 0,1, når BMD er defineret ved den relative definition.



**FIGUR 23.** Dækningsgrader af BMDL for fem forskellige metoder opdelt på størrelsen af BMR, 0,01, 0,05 og 0,1, når BMD er defineret ved "extra" definitionen.

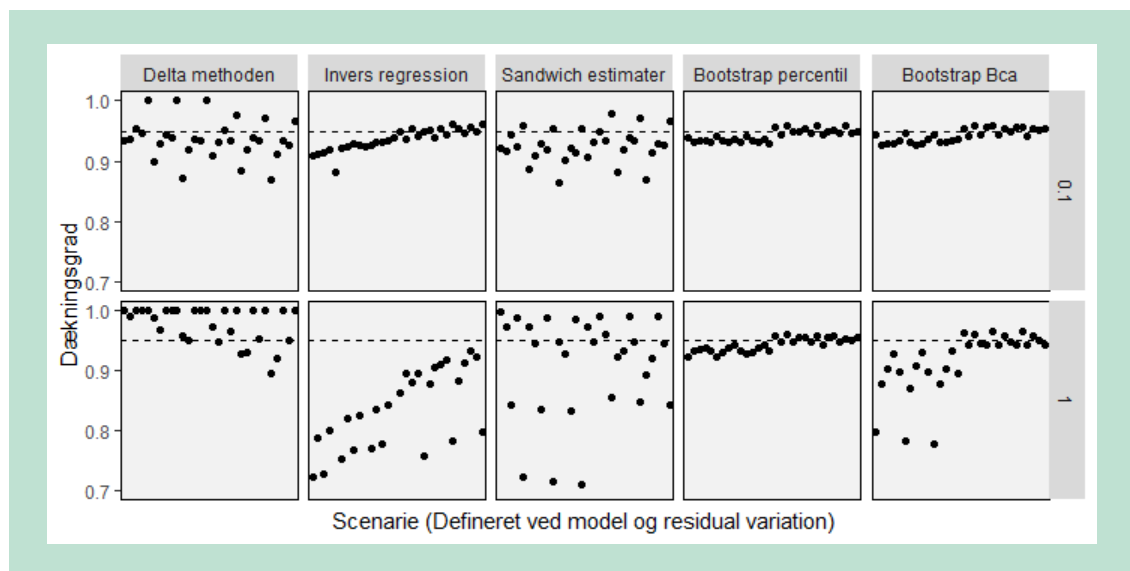


**FIGUR 24.** Dækningsgrader af BMDL for fem forskellige metoder opdelt på størrelsen af BMR, 0,01, 0,1 og 1,0, når BMD er defineret ved "added" definitionen.

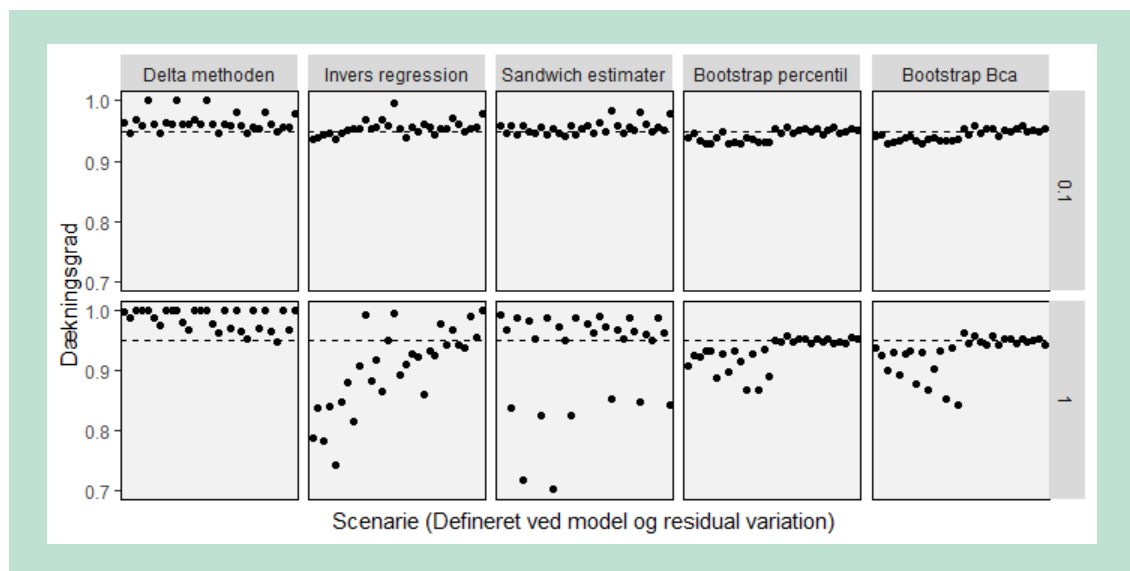
### 7.2.3 Variationen i data

Figur 25, 26 og 27 viser dækningsgrader af BMDL for de to forskellige niveauer af variation. Alle metoder klarede sig naturligt bedre for lave værdier af variation hvor kurven er bedre bestemt. Det gælt i særdeleshed deltametoden og invers regression hvad var forventeligt da begge bygger på asymptotiske metoder.

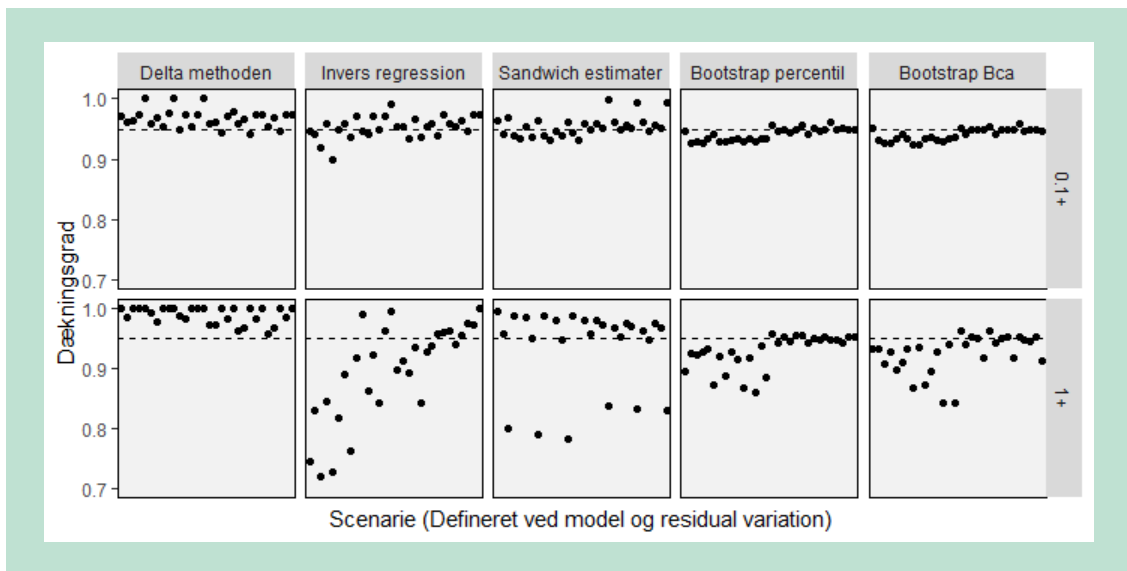
Figur 28 viser dækningsgrader for scenarier hvor der er tilføjet ekstra variation for nogle af doserne, dvs. for scenarier med varians heterogenitet. Umiddelbart ser det ikke ud som om nogle af de undersøgte metoder der påvirkes voldsomt af denne model-misspecifikation.



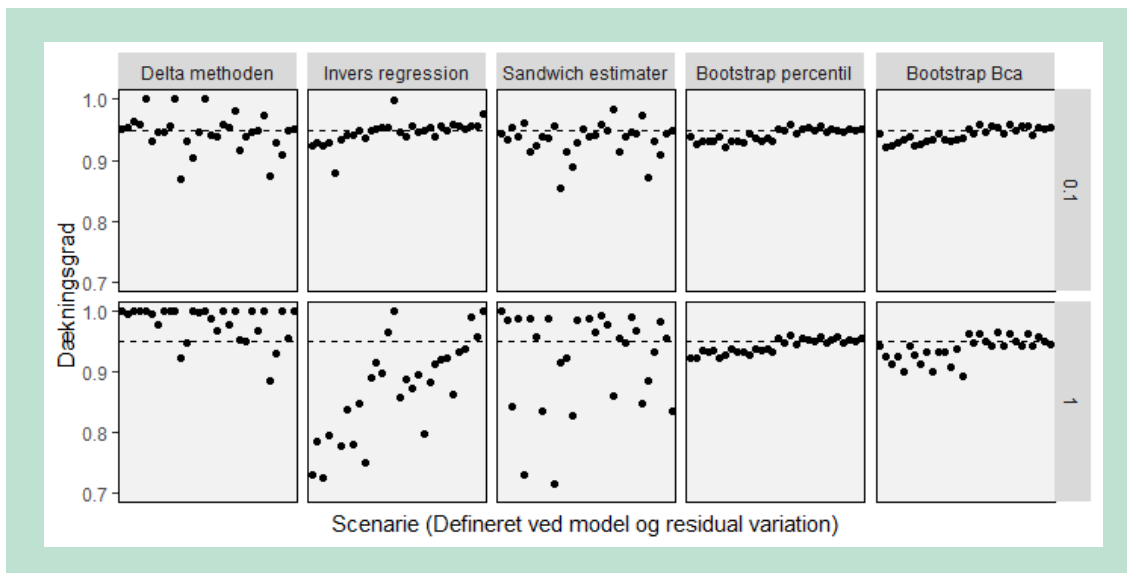
**FIGUR 25.** Dækningsgrader af BMDL for fem forskellige metoder opdelt på størrelsen af variationen i data, en standard afvigelse på 0,1 eller 1, når BMD er defineret ved den relative definition, når BMD er defineret ved den relative definition.



**FIGUR 26.** Dækningsgrader af BMDL for fem forskellige metoder opdelt på størrelsen af variationen i data, en standard afvigelse på 0,1 eller 1, når BMD er defineret ved "extra" definitionen, når BMD er defineret ved den relative definition, når BMD er defineret ved "extra" definitionen.

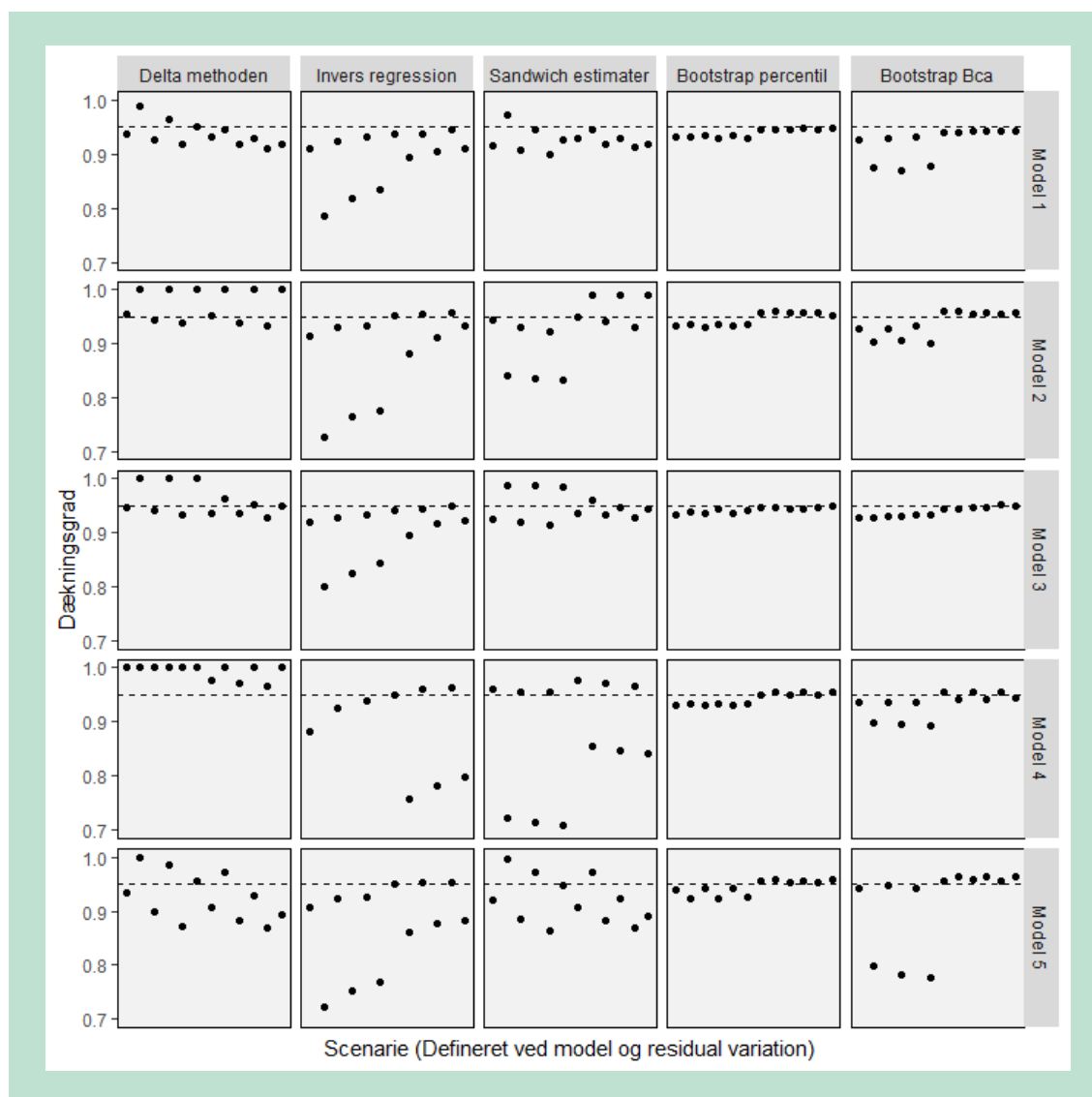


**FIGUR 27.** Dækningsgrader af BMDL for fem forskellige metoder opdelt på størrelsen af variationen i data, en gennemsnitlig standard afvigelse på 0,1 eller 1 med tilføjet ekstra varians inhomogenitet, når BMD er defineret ved "extra" definitionen.



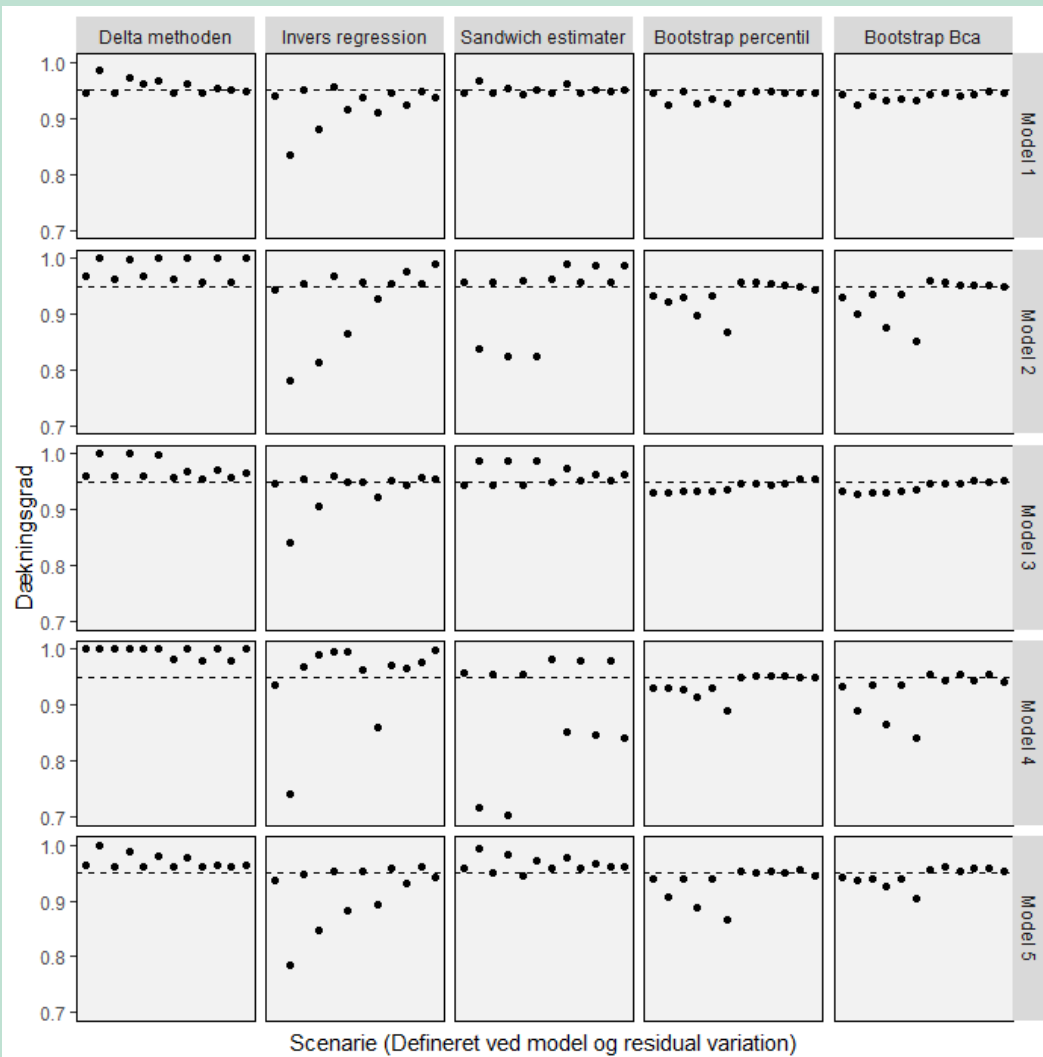
**FIGUR 28.** Dækningsgrader af BMDL for fem forskellige metoder opdelt på størrelsen af variationen i data, en standard afvigelse på 0,1 eller 1, når BMD er defineret ved "extra" definitionen, når BMD er defineret ved den relative definition, når BMD er defineret ved "added" definitionen.

## 7.2.4 Indflydelsen af den underliggende model

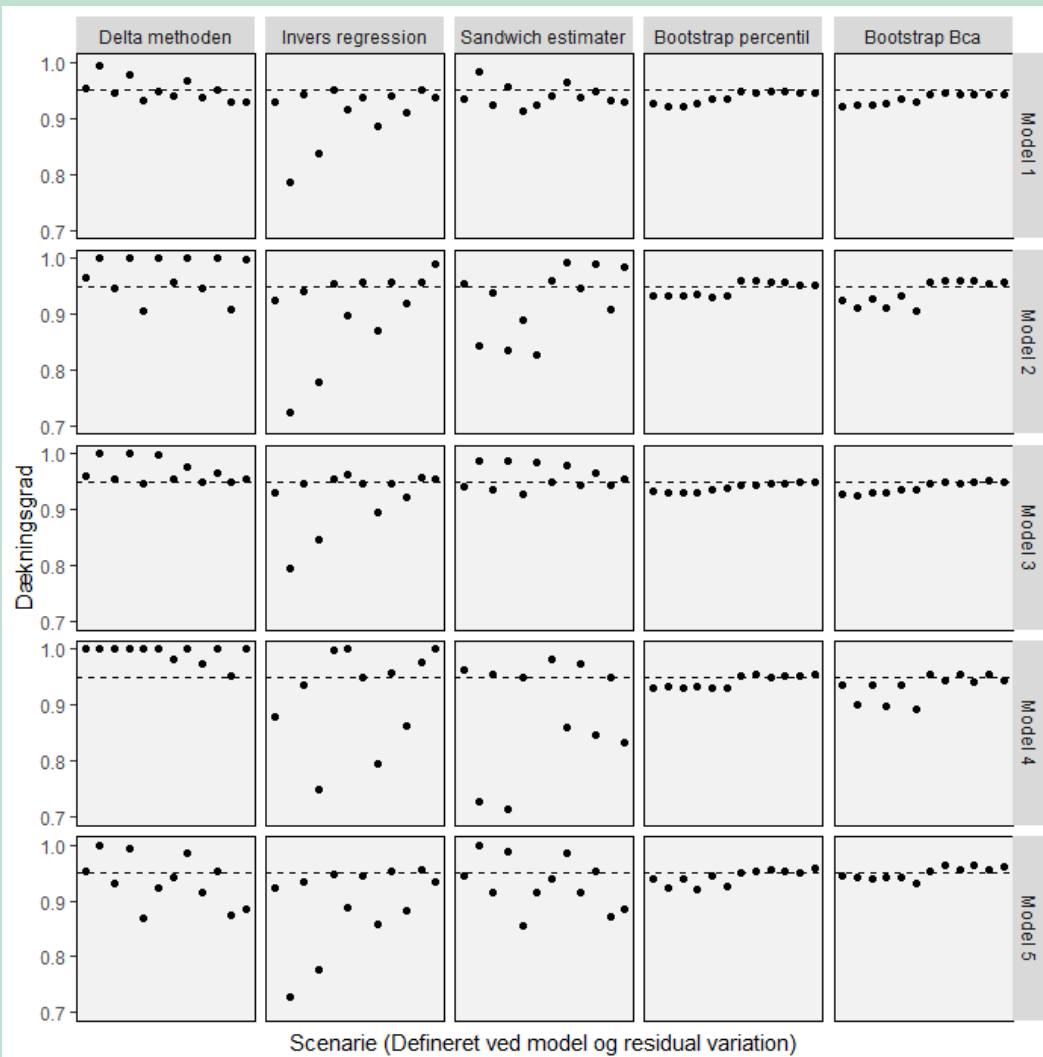


**FIGUR 29.** Dækningsgrader af BMDL for fem forskellige metoder opdelt på modellen brugt til at genere data, når BMD er defineret ved den relative definition.

Betragtes de underliggende modeller hver for sig ses det at model 1 og 3 generelt resulterede i de bedste dækningsgrader mens model 2 og 4 skilte sig ud i den anden retning med dæknings sandsynligheder længst fra det nominelle niveau (Figur 29, 30 og 31). Det kan skyldes at de førstnævnte modeller var godt dækket ind med doser til at beskrive hele kurven, mens de sidstnævnte modeller mangler data til at beskrive den øvre del af kurven. Denne forklaring passer også fint med at dette mønster er tydeligere ved brug af "extra" definitionen end den relative definition, da førstnævnte netop bruger øvre grænse i definitionen af BMD. Sandwich ramte rimelige dækningsgrader for model 1, 3 og 5 for "extra" definitionen, men i mindre grad for den relative definition.



**FIGUR 30.** Dækningsgrader af BMDL for fem forskellige metoder opdelt på modellen brugt til at genere data, når BMD er defineret ved "extra" definitionen.



**FIGUR 31.** Dækningsgrader af BMDL for fem forskellige metoder opdelt på modellen brugt til at genere data, når BMD er defineret ved "added" definitionen.

### 7.3 Opsummering

Baseret på et større simulationsstudie fandt vi at delta metoden leverer rimelige dækningsgrader, særligt for "extra" definitionen hvor den øvre grænse udnyttes i definitionen. Invers regression giver også rimelige dækningsgrader der modsat delta metoden dog lå under det nominelle niveau, særligt i scenarier med høj variation kombineret med få gentagelser.

BMDL baseret på sandwich estimatorer gav dårligere resultater end forventet. Det kan muligvis skyldes at de model-misspecifikationerne ikke har været udprægede nok i de kørte simulationer. Til gengæld viste bootstrap sig at klare sig overraskende godt i stort set alle scenarier. Bootstrap kan dermed anbefales generelt, og gerne den simple percentil-udgave der lidt overraskende overordnet set gav lidt bedre dækningsgrader end BCa, i hvert fald for den relative definition.

Generelt viser simulationerne, at det er klogt at have høje nok doser med til at kunne fange en øvre grænse af dosis-respons-sammenhængen. Særligt hvis en sådan bruges i definitionen af BMD.

## 8. Diskussion

Den seneste opdatering af EFSA's anbefalinger til brug af BMD metoden i risikovurdering slår fast at BMD foretrækkes frem for NOAEL tilgangen til at udlede DOD som udgangspunkt for grænseværdisættelse for kemikalier i fødevarer uanset kategori og oprindelse (Hardy et al. 2017). Der er dog ingen grund til at BMD ikke også kan benyttes indenfor økotoksikologien som bl.a. kommenteret af US EPA (Haber et al. 2018). Denne udmelding understøttes af nærværende rapport der gennem eksempler har illustreret anvendeligheden af BMD metoden på forskellige typer af økotoksikologiske data der ikke kun dækker de typiske dyrestudier brugt i human toksikologien. I det første af i alt seks eksempler, betragtedes det almindelige eksempel med binomiale data for en enkelt dosis-respons kurve. De næste to eksempler dækkede tælle- og kontinuerte data. For førstnævnte data type, blev det illustreret hvordan man kan håndtere forskellige udfordringer relateret til reproduktionsdata, en data type ofte brugt i økotoksikologien, men hvor håndteringen af tælledata er overset i de eksisterende software programmer udviklet til BMD bestemmelse. Eksempel fire illustrerede hvordan data fra mere komplekse design kan bruges til at bestemme BMD og BMDL ved en meta-analytisk tilgang der har vist sig at være brugbar for de relaterede ED/EC estimater (Xiaoqi Jiang and Kopp-Schneider 2014), mens det femte eksempel drejede sig om tid-til-event data. En data type der hidtil har været overset i BMD sammenhæng. Det sidste eksempel hvordan BMDL estimater potentielt kan indgå i SSD analyser. En analyse-type specifik for økotoksikologien.

EFSA anbefaler model-gennemsnit som den foretrukne metode til beregning af BMD og BMDL og efterspørger specifikt software der kan håndtere modelgennemsnit (Hardy et al. 2017). I mellemtiden er EFSA's software, PROAST, blevet opdateret med muligheden for modelgennemsnit (Slob 2018). For de kontinuerte data blev det i nærværende rapport illustreret hvordan den i projektet udviklede R-pakke tilbyder flere muligheder for at bestemme BMD og BMDL baseret på model-gennemsnit. Model-gennemsnit kan dog også bestemmes for andre typer af responsvariable.

Nærværende rapport har fokuseret på data fra eksperimentelle studier, men BMD metoden kan også benyttes for observationelle epidemiologiske studier (Hardy et al. 2017). For denne type data kan der dog være yderligere udfordringer idet data kan være uden kontroller, kræve justering for konfoundere og ofte kun dækker et begrænset interval af eksponeringskoncentrationer (Budtz-Jørgensen et al. 2013; Budtz-Jørgensen, Keiding, and Grandjean 2001; Hardy et al. 2017; Whitney and Ryan 2013).

Der er en klar tendens til, at mange af de i litteraturen foreslåede metoder resulterer i meget konservative BMDL. Selvom dette på overfladen kan virke fint ud fra et forsigtighedsprincip, bør målet og princippet være en realistisk risikovurdering.

I princippet kan BMD-metoden anvendes på alle datasæt der viser en dosis-respons trend, selv når der kun er få dosisgrupper til stede. I et sådant tilfælde skulle alle modeller, der potentiel kunne beskrive data, tages i betragtning, for eksempel ved at anvende modelgennemsnit. Konfidensintervaller for et estimeret BMD kan i tilfælde med meget få dosis-grupper dog være ekstremt brede (muligvis gående til uendelig) - men det afspejler den viden, der kan opnås fra data (Slob 2014a). I praksis bør et sådant tilfælde fra et statistisk synspunkt udløse nye eksperimenter og dermed flere data og måske revision af dosis valg (Hardy et al. 2017).

Som et alternativ til at generere nye data, vil en kombination af flere datasæt (inklusion af historiske data) muligvis kunne forbedre præcisionen hvormed BMD kan bestemmes. Relevante



data kunne i givet fald være data fra tidligere studier med samme test organisme, samme respons og fra samme laboratorie. Data der afviger på en eller flere af disse parametre kan måske også være relevant (Slob 2014a). Historiske data kan udnyttes på forskellige måder i BMD analyser: Data, nye og historiske, kan kombineres i en mixed model der tager højde for at data har forskellig oprindelse (som i eksempel 7.4 hvor data fra 9 uafhængige forsøg blev samlet); historiske data kan bruges til at bestemme baggrundssandsynligheden for en uønsket effekt; historiske data kan bruges til at forme priors i en bayesiansk dosis-respons analyse (Shao 2012); BMD-estimatet baseret på de nye data kan kombineres med BMD-estimer fra historiske data sæt i en meta-analytisk model (Ritz et al. 2019).

Metoderne foreslået her er dog ikke undersøgt til bunds i forhold til praktiske og analysemæssige udfordringer eller i hvor stor en grad indarbejdning af historiske data vil influere på konklusionerne for et givet datasæt.

Med den seneste opdatering af EFSA's guidelines for BMD analyser anbefales det kraftigt at genoverveje test guidelines så de tilpasses den forventede udbredelse af BMD metoden. De eksisterende guidelines er ikke optimale for BMD metoden der ville kunne drage nytte af mere frihed til at balancere mellem antallet af doser og gentagelser (Hardy et al. 2017). F.eks. har vi under simulationerne vist at det er en fordel at bruge en høj nok dosis til at fange en øvre grænse.

I dette projekt har vi valgt at implementere metoderne i det gratis dynamiske open source statistiske program R (R Core Team 2019). Denne tilgang tillader potentielle brugere fri adgang til de udviklede metoder i et i forvejen velbenyttet program til dosis-respons-modellering (Ritz et al. 2015), der gør R til et stort "udstillingsvindue" som kan sætte mere fokus på brug af BMD-metoden. I modsætning til de tidligere softwareinitiativer betyder udviklingen af open source-software også at projektets resultater i fremtiden kan forbedres eller videreudvikles af enhver der måtte være interesseret. Vi har allerede ovenfor nævnt nogle af de åbne spørgsmål, som det ville være naturligt at arbejde videre med.

## 9. Konklusion

BMD metoden kan fra et statistisk synspunkt anvendes indenfor økotoksikologi på samme vilkår som inden for human toksikologi, herunder som metode til at vurdere miljøeffekter i pesticidreguleringen. Der er udviklet en frit tilgængelig R-pakke der imødekommer flere af de problemstillinger der blev fundet ved en gennemgang af litteraturen indenfor BMD metoderne. Rapporten dokumenterer pakken der blandt andet giver mulighed for at finde BMDL ved bootstrap-metoden der i et simulationsstudie resulterede konsistent gode resultater (dækningsgrader meget tæt på det nominelle niveau). Projektet har desuden vist at BMD-metoden kan håndtere mange forskellige analyse-mæssige problemstillinger i økotoksikologien, også for andre typer af data end der hvor metoden allerede anvendes på i dag. Dette er illustreret gennem typiske data eksempler fra økotoksikologien. En nyskabelse som kan forbedre tolkningen af de klassiske SSD analyser er, at man nu kan benytte BMDL i stedet for ED<sub>50</sub> eller NOAEL eller LOAEL.

# 10. Litteratur

- Al-Saidy, Obaid M., Walter W. Piegorsch, R. Webster West, and Daniela K. Nitscheva. 2003. "Confidence Bands for Low-Dose Risk Estimation with Quantal Response Data." *Biometrics* 59(4): 1056–62.
- Bailer, A. John, Robert B. Noble, and Matthew W. Wheeler. 2005. "Model Uncertainty and Risk Estimation for Experimental Studies of Quantal Responses." *Risk Analysis* 25(2): 291–99.
- Bailer, A. John, and James T. Oris. 1997. "Estimating Inhibition Concentrations for Different Response Scales Using Generalized Linear Models." *Environmental Toxicology and Chemistry* 16(7): 1554–59.
- . 2000. "Defining the Baseline for Inhibition Concentration Calculations for Hormetic Hazards." *Journal of Applied Toxicology* 20(2): 121–25.
- Begon, Michael, Colin R. Townsend, and John L. Harper. 2005. *Ecology: From Individuals to Ecosystems*. 4th ed.
- Bresson, Jean-Louis et al. 2009. "Scientific Opinion of the Panel on Dietetic Products, Nutrition and Allergies." *The EFSA Journal* 999(December): 1–9.
- Buckland, S. T., K. P. Burnham, and N. H. Augustin. 1997. "Model Selection: An Integral Part of Inference." *Biometrics* 53(2): 603–18.
- Buckley, Brooke E, and Walter W Piegorsch. 2008. "Simultaneous Confidence Bands for Abbott-Adjusted Quantal Response Models." *Statistical methodology* 5(3): 209–19.
- Buckley, Brooke E, Walter W Piegorsch, and R Webster West. 2009. "Confidence Limits on One-Stage Model Parameters in Benchmark Risk Assessment." *Environmental and ecological statistics* 16(1): 53–62.
- Budtz-Jørgensen, Esben. 2007. "Estimation of the Benchmark Dose by Structural Equation Models." *Biostatistics* 8(4): 675–88.
- Budtz-Jørgensen, Esben, David Bellinger, Bruce Lanphear, and Philippe Grandjean. 2013. "An International Pooled Analysis for Obtaining a Benchmark Dose for Environmental Lead Exposure in Children." *Risk Analysis* 33(3): 450–61.
- Budtz-Jørgensen, Esben, Niels Keiding, and Philippe Grandjean. 2001. "Benchmark Dose Calculation from Epidemiological Data." *Biometrics* 57(3): 698–706.
- Burgoon, Lyle D., and Timothy R. Zacharewski. 2008. "Automated Quantitative Dose-Response Modeling and Point of Departure Determination for Large Toxicogenomic and High-Throughput Screening Data Sets." *Toxicological Sciences* 104(2): 412–18.
- Carr, Gregory J., A. John Bailer, Jane M. Rawlings, and Scott E. Belanger. 2018. "On the Impact of Sample Size on Median Lethal Concentration Estimation in Acute Fish Toxicity Testing: Is  $n = 7/\text{Group}$  Enough?" *Environmental Toxicology and Chemistry* 37(6): 1565–78.
- Cedergreen, Nina et al. 2017. "Can Toxicokinetic and Toxicodynamic Modeling Be Used to Understand and Predict Synergistic Interactions between Chemicals?" *Environmental Science & Technology* 51(24): 14379–89.
- Cedergreen, Nina, Nils Jakob Nørhøve, Claus Svendsen, and David J. Spurgeon. 2016. "Variable Temperature Stress in the Nematode *Caenorhabditis Elegans* (Maupas) and Its Implications for Sensitivity to an Additional Chemical Stressor" ed. Jonathan H Freedman. *PLOS ONE* 11(1): e0140277.
- Chen, Chu-Chih, and James J. Chen. 2014. "Benchmark Dose Calculation for Ordered Categorical Responses." *Risk Analysis* 34(8): 1435–47.
- Crane, Mark, Michael C Newman, Peter F Chapman, and John S Fenlon. 2016. *Risk Assessment with Time to Event Models*. CRC Press.
- Crump, Kenny S. 1984. "A New Method for Determining Allowable Daily Intakes." *Fundamental and Applied Toxicology* 4(5): 854–71.
- . 1995. "Calculation of Benchmark Doses from Continuous Data." *Risk Analysis* 15(1): 79–89.
- Crump, Kenny S et al. 2010. "What Role for Biologically Based Dose-Response Models in Estimating Low-Dose Risk?" *Environmental health perspectives* 118(5): 585–88.

- Das, Ishapathik. 2018. "Robust Benchmark Dose Estimation Using an Infinite Family of Response Functions." *Human and Ecological Risk Assessment: An International Journal* 24(8): 2054–69.
- Davis, J. Allen, Jeffrey S. Gift, and Q. Jay Zhao. 2011. "Introduction to Benchmark Dose Methods and U.S. EPA's Benchmark Dose Software (BMDS) Version 2.1.1." *Toxicology and Applied Pharmacology* 254(2): 181–91.
- Davison, A. C., and D. V. Hinkley. 2007. *Bootstrap Methods and Their Application*. 7th ed. New York: Cambridge University Press.
- Delignette-Muller, Marie Laure, Christelle Lopes, Philippe Veber, and Sandrine Charles. 2014. "Statistical Handling of Reproduction Data for Exposure-Response Modeling." *Environmental Science and Technology* 48: 7544–51.
- Demidenko, Eugene, Benjamin B Williams, Ann Barry Flood, and Harold M Swartz. 2013. "Standard Error of Inverse Prediction for Dose-Response Relationship: Approximate and Exact Statistical Inference." *Statistics in medicine* 32(12): 2048–61.
- Deutsch, Roland C., and Walter W. Piegorsch. 2012. "Benchmark Dose Profiles for Joint-Action Quantal Data in Quantitative Risk Assessment." *Biometrics* 68(4): 1313–22.
- . 2013. "Benchmark Dose Profiles for Joint-Action Continuous Data in Quantitative Risk Assessment." *Biometrical Journal* 55(5): 741–54.
- Edler, L, K Poirier, M Dourson, and J Kleiner. 2002. "Mathematical Modelling and Quantitative Methods." *Food and Chemical* 40: 283–326.
- European Chemicals Agency. 2017. *Guidance on Biocidal Products Regulation: Volume III Human Health - Assessment & Evaluation (Parts B+C)*.
- European Food Safety Authority scientific committee. 2009. "Guidance of the Scientific Committee on Use of the Benchmark Dose Approach in Risk Assessment." *EFSA Journal* 1150: 1–72.
- . 2016. "GUIDANCE DOCUMENT Use of the Benchmark Dose Approach in Risk Assessment BMD Approach in Risk Assessment." : 1–51.
- Faes, Christel, Marc Aerts, Helena Geys, and Geert Molenberghs. 2007. "Model Averaging Using Fractional Polynomials to Estimate a Safe Level of Exposure." *Risk Analysis* 27(1): 111–23.
- Fang, Q. et al. 2015. "Bayesian Model-Averaged Benchmark Dose Analysis via Reparameterized Quantal-Response Models." *Biometrics* 71(4): 1168–75.
- Fang, Q., W. W. Piegorsch, and K. Y. Barnes. 2015. "Bayesian Benchmark Dose Analysis." *Environmetrics* 26(5): 373–82.
- Farmahin, Reza et al. 2017. "Recommended Approaches in the Application of Toxicogenomics to Derive Points of Departure for Chemical Risk Assessment." *Archives of Toxicology* 91(5): 2045–65.
- Fox, John F., Karen A. Hogan, and Allen Davis. 2017. "Dose-Response Modeling with Summary Data from Developmental Toxicity Studies." *Risk Analysis* 37(5): 905–17.
- Gaylor, David W., and James J. Chen. 1996. "Precision of Benchmark Dose Estimates for Continuous (Nonquantal) Measurements of Toxic Effects." *Regulatory Toxicology and Pharmacology* 24(1): 19–23.
- Gaylor, David W., and W Slikker. 1990. "Risk Assessment for Neurotoxic Effects." *Neurotoxicology* 11(2): 211–18.
- Gaylor, David W., and William Slikker. 2004. "Role of the Standard Deviation in the Estimation of Benchmark Doses with Continuous Data." *Risk Analysis* 24(6): 1683–87.
- Gottardi, Michele, and Nina Cedergreen. 2019. "The Synergistic Potential of Azole Fungicides Does Not Directly Correlate to the Inhibition of Cytochrome P450 Activity in Aquatic Invertebrates." *Aquatic Toxicology* 207: 187–96.
- Guha, Nilabja et al. 2013. "Nonparametric Bayesian Methods for Benchmark Dose Estimation." *Risk Analysis* 33(9): 1608–19.
- Haber, Lynne T. et al. 2018. "Benchmark Dose (BMD) Modeling: Current Practice, Issues, and Challenges." *Critical Reviews in Toxicology* 48(5): 387–415.
- Hardy, Anthony et al. 2017. "Update: Use of the Benchmark Dose Approach in Risk Assessment." *EFSA Journal* 15(1): 1–41.
- Hoekstra, J. A. 1987. "Acute Bioassays with Control Mortality." *Water, Air, and Soil Pollution* 35(3–4): 311–17.
- Hughes, Michael R., A. John Bailer, and Debra L. Denton. 2001. "Toxicant- and Response-

- Specific Comparisons of Statistical Methods for Estimating Effective Concentrations.” *Environmental Toxicology and Chemistry* 20(6): 1374–80.
- International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. 2014. “0013-Guideline for Elemental Impurities Q3D.” *ICH* (December): 1–75.
- International programme on chemical safety, World Health Organization. 2009. *Principles and Methods for the Risk Assessment of Chemicals in Food*. World Health Organization.
- Izadi, Hoda, Jean E. Grundy, and Ranjan Bose. 2012. “Evaluation of the Benchmark Dose for Point of Departure Determination for a Variety of Chemical Classes in Applied Regulatory Settings.” *Risk Analysis* 32(5): 830–35.
- Jensen, Signe M., Felix M. Kluxen, and Christian Ritz. 2019. “A Review of Recent Advances in Benchmark Dose Methodology.” *Risk Analysis*: risa.13324.
- Jensen, Signe M., and Christian Ritz. 2015. “Simultaneous Inference for Model Averaging of Derived Parameters.” *Risk Analysis* 35(1): 68–76.
- Jiang, Xiaogang, Hans Chr Bruun Hansen, Bjarne W. Strobel, and Nina Cedergreen. 2018. “What Is the Aquatic Toxicity of Saponin-Rich Plant Extracts Used as Biopesticides?” *Environmental Pollution* 236: 416–24.
- Jiang, Xiaoqi, and Annette Kopp-Schneider. 2014. “Summarizing EC50 Estimates from Multiple Dose-Response Experiments: A Comparison of a Meta-Analysis Strategy to a Mixed-Effects Model Approach.” *Biometrical Journal* 56(3): 493–512.
- Kang, Seung-Ho, Ralph L. Kodell, and James J. Chen. 2000. “Incorporating Model Uncertainties along with Data Uncertainties in Microbial Risk Assessment.” *Regulatory Toxicology and Pharmacology* 32(1): 68–72.
- Kerns, Lucy X. 2017. “Simultaneous Confidence Bands for Log-Logistic Regression with Applications in Risk Assessment.” *Biometrical Journal* 59(3): 420–29.
- . 2018. “Multiplicity-Adjusted Confidence Limits in Risk Assessment with Quantal Response Data.” *Journal of Biopharmaceutical Statistics* 28(6): 1182–92.
- Kim, Steven B, Ralph L Kodell, and Hojin Moon. 2014. “A Diversity Index for Model Space Selection in the Estimation of Benchmark and Infectious Doses via Model Averaging.” *Risk analysis: an official publication of the Society for Risk Analysis* 34(3): 453–64.
- Kodell, Ralph L., and Ronnie W. West. 1993. “Upper Confidence Limits on Excess Risk for Quantitative Responses.” *Risk Analysis* 13(2): 177–82.
- Kopylev, Leonid, and John Fox. 2009. “Parameters of a Dose-Response Model Are on the Boundary: What Happens with BMDL?” *Risk Analysis* 29(1): 18–25.
- Lin, Lizhen, Walter W. Piegorsch, and Rabi Bhattacharya. 2015. “Nonparametric Benchmark Dose Estimation with Continuous Dose-Response Data.” *Scandinavian Journal of Statistics* 42(3): 713–31.
- Liu, Jingyu, Walter W. Piegorsch, A. Grant Schissler, and Susan L. Cutter. 2018. “Autologistic Models for Benchmark Risk or Vulnerability Assessment of Urban Terrorism Outcomes.” *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 181(3): 803–23.
- MacGregor, James T. et al. 2015. “IWGT Report on Quantitative Approaches to Genotoxicity Risk Assessment I. Methods and Metrics for Defining Exposure-Response Relationships and Points of Departure (PoDs).” *Mutation Research/Genetic Toxicology and Environmental Mutagenesis* 783: 55–65.
- Mcculloch, Charles E., and John M. Neuhaus. 2001. *Generalized Linear Mixed Models*. Chichester, UK: John Wiley & Sons, Ltd.
- Moerbeek, Mirjam, Aldert H. Piersma, and Wout Slob. 2004. “A Comparison of Three Methods for Calculating Confidence Intervals for the Benchmark Dose.” *Risk Analysis* 24(1): 31–40.
- Moon, Hojin et al. 2013. “Model Uncertainty and Model Averaging in the Estimation of Infectious Doses for Microbial Pathogens.” *Risk Analysis* 33(2): 220–31.
- Moon, Hojin, Hyun-Joo Kim, James J. Chen, and Ralph L. Kodell. 2005. “Model Averaging Using the Kullback Information Criterion in Estimating Effective Doses for Microbial Infection and Illness.” *Risk Analysis* 25(5): 1147–59.
- Morales, Knashawn H., and Louise M. Ryan. 2005. “Benchmark Dose Estimation Based on Epidemiologic Cohort Data.” *Environmetrics* 16(5): 435–47.
- Muri, Stefan D., Josef R. Schlatter, and Beat J. Brüschiweiler. 2009. “The Benchmark Dose Approach in Food Risk Assessment: Is It Applicable and Worthwhile?” *Food and*

- Chemical Toxicology* 47(12): 2906–25.
- Najita, Julie S., and Paul J. Catalano. 2013. "On Determining the BMD from Multiple Outcomes in Developmental Toxicity Studies When One Outcome Is Intentionally Missing." *Risk Analysis* 33(8): 1500–1509.
- Najita, Julie S., Yi Li, and Paul J. Catalano. 2009. "A Novel Application of a Bivariate Regression Model for Binary and Continuous Outcomes to Studies of Fetal Toxicity." *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 58(4): 555–73.
- Namata, Harriet, Marc Aerts, Christel Faes, and Peter Teunis. 2008. "Model Averaging in Microbial Risk Assessment Using Fractional Polynomials." *Risk Analysis* 28(4): 891–905.
- Newman, Michael C., and Philip M. Dixon. 1996. "Ecologically Meaningful Estimates of Lethal Effect in Individuals." In *Ecotoxicology: A Hierarchical Treatment*, eds. Michael C. Newman and Charles H. Jagoe. Boca Raton, Florida: Lewis Publishers, CRC, 225–53.
- Nitcheva, Daniela K., Walter W. Piegorsch, R. Webster West, and Ralph L. Kodell. 2005. "Multiplicity-Adjusted Inferences in Risk Assessment: Benchmark Analysis with Quantal Response Data." *Biometrics* 61(1): 277–86.
- Noble, Robert B., A. John Bailer, and Robert Park. 2009. "Model-Averaged Benchmark Concentration Estimates for Continuous Response Data Arising from Epidemiological Studies." *Risk Analysis* 29(4): 558–64.
- Öberg, Mattias. 2010. "Benchmark Dose Approaches in Chemical Health Risk Assessment in Relation to Number and Distress of Laboratory Animals." *Regulatory Toxicology and Pharmacology* 58(3): 451–54.
- OECD. 2014. *Guidance Document 116 on the Conduct and Design of Chronic Toxicity and Carcinogenicity Studies, Supporting Test Guidelines 451, 452 and 453*. OECD Publishing.
- Organisation for Economic Co-operation and Development. 1998a. *Test No. 408: Repeated Dose 90-Day Oral Toxicity Study in Rodents*. OECD Publishing.
- . 1998b. *Test No. 409: Repeated Dose 90-Day Oral Toxicity Study in Non-Rodents*. OECD Publishing.
- . 2001. *Test No. 416: Two-Generation Reproduction Toxicity*. OECD Publishing.
- . 2008. *Test No. 407: Repeated Dose 28-Day Oral Toxicity Study in Rodents*. OECD Publishing.
- . 2009a. *Test No. 451: Carcinogenicity Studies*. OECD Publishing, Paris.
- . 2009b. *Test No. 452: Chronic Toxicity Studies*. OECD Publishing, Paris.
- . 2009c. *Test No. 453: Combined Chronic Toxicity/Carcinogenicity Studies*. OECD Publishing.
- . 2012. *Guidance Document 116 on the Conduct and Design of Chronic Toxicity and Carcinogenicity Studies, Supporting Test Guidelines 451, 452 and 453*. Second edi. Paris.
- . 2014. *Current Approaches in the Statistical Analysis of Ecotoxicity Data*. OECD Publishing.
- . 2017a. *Test No. 412: Subacute Inhalation Toxicity: 28-Day Study*. OECD Publishing.
- . 2017b. *Test No. 413: Subchronic Inhalation Toxicity: 90-Day Study*. OECD Publishing.
- Organisation for Economic Co-operation and Development (OECD). 2006. *Current Approaches in the Statistical Analysis of Ecotoxicity Data: A Guidance to Application - Annexes*.
- Peña, Edsel A. et al. 2017. "Model Selection and Estimation with Quantal-Response Data in Benchmark Risk Assessment." *Risk Analysis* 37(4): 716–32.
- Peng, Jianan, Megan Robichaud, and Abdelaziz Q. Alsubie. 2015. "Simultaneous Confidence Bands for Low-Dose Risk Estimation with Quantal Data." *Biometrical Journal* 57(1): 27–38.
- Piegorsch, Walter W. et al. 2013. "Information-Theoretic Model-Averaged Benchmark Dose Analysis in Environmental Risk Assessment." *Environmetrics* 24(3): 143–57.
- Piegorsch, Walter W., Hui Xiong, Rabi N. Bhattacharya, and Lizhen Lin. 2012. "Nonparametric Estimation of Benchmark Doses in Environmental Risk Assessment." *Environmetrics* 23(8): 717–28.
- . 2014. "Benchmark Dose Analysis via Nonparametric Regression Modeling." *Risk Analysis* 34(1): 135–51.

- Posthuma, Leo, Glenn W Suter II, and Theo P Traas. 2002. *Species Sensitivity Distributions in Ecotoxicology*. First. ed. Michael C. Newman. Boca Raton, Florida: Lewis Publishers, CRC Press.
- R Core Team. 2019. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.r-project.org/>.
- Ringblom, Joakim, Gunnar Johanson, and Mattias Öberg. 2014. "Current Modeling Practice May Lead to Falsely High Benchmark Dose Estimates." *Regulatory Toxicology and Pharmacology* 69(2): 171–77.
- Ringblom, Joakim, Fereshteh Kalantari, Gunnar Johanson, and Mattias Öberg. 2017. "Influence of Distribution of Animals between Dose Groups on Estimated Benchmark Dose and Animal Welfare for Continuous Effects." *Risk Analysis* 37(9): 1716–28.
- Risk Assessment Forum, and U.S. Environmental Protection Agency. 2005. "Guidelines for Carcinogen Risk Assessment." : 1–166.
- Ritz, Christian, Florent Baty, Jens C Streibig, and Daniel Gerhard. 2015. "Dose-Response Analysis Using R." *PLoS one* 10(12): e0146021.
- Ritz, Christian, Daniel Gerhard, and Ludwig A. Hothorn. 2013. "A Unified Framework for Benchmark Dose Estimation Applied to Mixed Models and Model Averaging." *Statistics in Biopharmaceutical Research* 5(1): 79–90.
- Ritz, Christian, Signe Marie Jensen, Daniel Gerhard, and Jens Carl Streibig. 2019. *Dose-Response Analysis Using R*. 1st ed. Boca Raton, Florida: Chapman & Hall.
- Roberts, Steven, and Michael A Martin. 2010. "Bootstrap-after-Bootstrap Model Averaging for Reducing Model Uncertainty in Model Selection for Air Pollution Mortality Studies." *Environmental health perspectives* 118(1): 131–36.
- Sand, Salomon, Dietrich von Rosen, Katarina Victorin, and Agneta Falk Filipsson. 2006. "Identification of a Critical Dose Level for Risk Assessment: Developments in Benchmark Dose Analysis of Continuous Endpoints." *Toxicological Sciences* 90(1): 241–51.
- Scientific Committee on Consumer Safety. 2016. 1564 *The SCCS's Notes of Guidance for the Testing of Cosmetic Ingredients*.
- Shao, Kan. 2012. "A Comparison of Three Methods for Integrating Historical Information for Bayesian Model Averaged Benchmark Dose Estimation." *Environmental Toxicology and Pharmacology* 34(2): 288–96.
- Shao, Kan, Jeffrey S. Gift, and R. Woodrow Setzer. 2013. "Is the Assumption of Normality or Log-Normality for Continuous Response Data Critical for Benchmark Dose Estimation?" *Toxicology and Applied Pharmacology* 272(3): 767–79.
- Shao, Kan, and Mitchell J. Small. 2011. "Potential Uncertainty Reduction in Model-Averaged Benchmark Dose Estimates Informed by an Additional Dose Study." *Risk Analysis* 31(10): 1561–75.
- . 2012. "Statistical Evaluation of Toxicological Experimental Design for Bayesian Model Averaged Benchmark Dose Estimation with Dichotomous Data." *Human and Ecological Risk Assessment: An International Journal* 18(5): 1096–1119.
- Simmons, Susan J. et al. 2015. "Bayesian Model Averaging for Benchmark Dose Estimation." *Environmental and Ecological Statistics* 22(1): 5–16.
- Slikker, William et al. 2004. "Dose-Dependent Transitions in Mechanisms of Toxicity." *Toxicology and Applied Pharmacology* 201(3): 203–25.
- Slob, Wout. 2002. "Dose-Response Modeling of Continuous Endpoints." *Toxicological Sciences* 66(2): 298–312.
- . 2014a. "Benchmark Dose and the Three Rs. Part I. Getting More Information from the Same Number of Animals." *Critical Reviews in Toxicology* 44(7): 557–67.
- . 2014b. "Benchmark Dose and the Three Rs. Part II. Consequences for Study Design and Animal Use." *Critical Reviews in Toxicology* 44(7): 568–80.
- . 2017. "A General Theory of Effect Size, and Its Consequences for Defining the Benchmark Response (BMR) for Continuous Endpoints." *Critical Reviews in Toxicology* 47(4): 342–51.
- . 2018. "Joint Project on Benchmark Dose Modelling with RIVM." *EFSA Supporting Publications* 15(12): 1–14.
- Slob, Wout, Mirjam Moerbeek, Eija Rauniomaa, and Aldert H. Piersma. 2005. "A Statistical Evaluation of Toxicity Study Designs for the Estimation of the Benchmark Dose in

- Continuous Endpoints." *Toxicological Sciences* 84(1): 167–85.
- Slob, Wout, and M. N. Pieters. 1998. "A Probabilistic Approach for Deriving Acceptable Human Intake Limits and Human Health Risks from Toxicological Studies: General Framework." *Risk Analysis* 18(6): 787–98.
- Slob, Wout, and R. Woodrow Setzer. 2014. "Shape and Steepness of Toxicological Dose–Response Relationships of Continuous Endpoints." *Critical Reviews in Toxicology* 44(3): 270–97.
- Smetanová, Soňa et al. 2015. "High-Throughput Concentration-Response Analysis for Omics Datasets." *Environmental Toxicology and Chemistry* 34(9): 2167–80.
- Taft, Sarah C., and Stephanie A. Hines. 2012. "Benchmark Dose Analysis for *Bacillus Anthracis* Inhalation Exposures in the Nonhuman Primate." *Risk Analysis* 32(10): 1750–68.
- Thomas, Russell S. et al. 2007. "A Method to Integrate Benchmark Dose Estimates with Genomic Data to Assess the Functional Effects of Chemical Exposure." *Toxicological Sciences* 98(1): 240–48.
- . 2011. "Application of Transcriptional Benchmark Dose Values in Quantitative Cancer and Noncancer Risk Assessment." *Toxicological Sciences* 120(1): 194–205.
- US Environmental Protection Agency. 2012. *Benchmark Dose Technical Guidance*. Washington, DC.
- Varewyck, Machteld, and Tobias Verbeke. 2017. "Software for Benchmark Dose Modelling." *EFSA Supporting Publications* 14(2): 1–15.
- Viechtbauer, Wolfgang. 2010. "Conducting Meta-Analyses in R with the Metafor Package." *Journal of Statistical Software* 36(3): 1–48.
- West, R. Webster et al. 2012. "The Impact of Model Uncertainty on Benchmark Dose Estimation." *Environmetrics* 23(8): 706–16.
- West, R. Webster, Daniela K. Nitcheva, and Walter W. Piegorsch. 2009. "Bootstrap Methods for Simultaneous Benchmark Analysis with Quantal Response Data." *Environmental and Ecological Statistics* 16(1): 63–73.
- Wheeler, Matthew W. et al. 2017. "Bayesian Quantile Impairment Threshold Benchmark Dose Estimation for Continuous Endpoints." *Risk Analysis* 37(11): 2107–18.
- Wheeler, Matthew W., and A. John Bailer. 2007. "Properties of Model-Averaged BMDLs: A Study of Model Averaging in Dichotomous Response Risk Estimation." *Risk Analysis* 27(3): 659–70.
- . 2008. "Model Averaging Software for Dichotomous Dose Response Risk Estimation." *Journal of Statistical Software* 26(5): 1–15.
- . 2009a. "Comparing Model Averaging with Other Model Selection Strategies for Benchmark Dose Estimation." *Environmental and Ecological Statistics* 16(1): 37–51.
- . 2012. "Monotonic Bayesian Semiparametric Benchmark Dose Analysis." *Risk Analysis* 32(7): 1207–18.
- . 2013. "An Empirical Comparison of Low-Dose Extrapolation from Points of Departure (PoD) Compared to Extrapolations Based upon Methods That Account for Model Uncertainty." *Regulatory Toxicology and Pharmacology* 67(1): 75–82.
- Wheeler, Matthew W., and A. John Bailer. 2009b. "Benchmark Dose Estimation Incorporating Multiple Data Sources." *Risk Analysis* 29(2): 249–56.
- Wheeler, Matthew W., R. M. Park, A. J. Bailer, and C. Whittaker. 2015. "Historical Context and Recent Advances in Exposure-Response Estimation for Deriving Occupational Exposure Limits." *Journal of Occupational and Environmental Hygiene* 12(sup1): S7–17.
- Wheeler, Matthew W., Kan Shao, and A. John Bailer. 2015. "Quantile Benchmark Dose Estimation for Continuous Endpoints." *Environmetrics* 26(5): 363–72.
- Whitney, Melissa, and Louise Ryan. 2013. "Uncertainty Due to Low-dose Extrapolation: Modified BMD Methodology for Epidemiological Data." *Environmetrics* 24(5): 289–97.
- World Health Organization. 2014. "Harmonization Project Document 11 Guidance Document on Evaluating and Expressing Uncertainty in Hazard Characterization." *World Health Organization*: 1–158.
- Wu, Yuping et al. 2006. "Multiplicity-Adjusted Inferences in Risk Assessment: Benchmark Analysis with Continuous Response Data." *Environmental and Ecological Statistics* 13(1): 125–41.
- Yang, Longlong, Bruce C Allen, and Russell S Thomas. 2007. "BMDEpress: A Software Tool



- for the Benchmark Dose Analyses of Genomic Data." *BMC Genomics* 8(1): 387.
- Yu, Zi-Fan, and Paul J. Catalano. 2005. "Quantitative Risk Assessment for Multivariate Continuous Outcomes with Application to Neurotoxicology: The Bivariate Case." *Biometrics* 61(3): 757–66.
- Zeller, Andreas, Gonzalo Duran-Pacheco, and Melanie Guérard. 2017. "An Appraisal of Critical Effect Sizes for the Benchmark Dose Approach to Assess Dose–Response Relationships in Genetic Toxicology." *Archives of Toxicology* 91(12): 3799–3807.
- Zhang, Jing, A. John Bailer, and James T. Oris. 2012a. "Bayesian Approach to Estimating Reproductive Inhibition Potency in Aquatic Toxicity Testing." *Environmental Toxicology and Chemistry* 31(4): 916–27.
- . 2012b. "Bayesian Approach to Potency Estimation for Aquatic Toxicology Experiments When a Toxicant Affects Both Fecundity and Survival." *Environmental Toxicology and Chemistry* 31(8): 1920–30.
- Zhang, Jing, A. John Bailer, and James T. Oris. 2014. "Estimating Potency for Hierarchical Dichotomous Responses in an Aquatic Toxicology Study." 19(2): 185–201.
- Zhu, Yiliang et al. 2005. "Analyses of Neurobehavioral Screening Data: Benchmark Dose Estimation." *Regulatory Toxicology and Pharmacology* 42(2): 190–201.
- Zhu, Yiliang, Tao Wang, and Jenny Z.H. Jelsovsky. 2007. "Bootstrap Estimation of Benchmark Doses and Confidence Limits with Clustered Quantal Data." *Risk Analysis* 27(2): 447–65.

# Bilag 1.

## Bilag 1.1 Forkortelser

AIC	Akaïke information criteria
BMD	Benchmark dose
BMDL	Benchmark dose nedre grænse
BMR	Benchmark response
DOD	Dose of departure
EC	Effective concentration
ED	Effective dose
EFSA	European food safety authority
LC	Lethal concentration
LD	Lethal dose
LOAEL	Lowest Observed Adverse Effect Level
NOAEL	No Observed Adverse Effect Level
OECD	Organisationen for økonomisk samarbejde og udvikling
US EPA	United States Environmental Protection Agency

## Benchmark dose – Bedre udnyttelse af data i pesticidregistrering og –regulering

Formålet med dette projekt har været at forbedre vidensgrundlaget for den aktuelle pesticid-regulering ved at introducere benchmark dose (BMD) metoden for en mere datadrevet og objektiv vurdering med en dertil hørende potentiel reduktion i miljørisiko og påvirkning med pesticider.

BMD metoden bruges til på baggrund af en dosis-respons model at finde doser som udgangspunkt for en farlighedsvurdering i risikovurderingsprocessen. BMD er defineret som den dosis, der resulterer i en lille på forhånd fastlagt "acceptabel" stigning i sandsynligheden for et uønsket event i forhold til baggrundssandsynligheden. I praksis bruges ofte den nedre grænse for konfidensintervallet for den estimerede BMD (BMDL) som udgangspunkt for grænseværdifastsættelse.

BMD metoden har stort potentiale indenfor økotoxikologi, herunder som metode til at vurdere miljøeffekter i pesticidreguleringen. Der er udviklet en frit tilgængelig R-pakke der imødekommer flere af de problemstillinger der blev fundet ved en gennemgang af litteraturen indenfor BMD metoderne.

Denne rapport dokumenterer pakken der blandt andet giver mulighed for at finde BMDL ved bootstrap-metoden der i et simulationsstudie resulterede i konsistent gode resultater (dækningsgrader meget tæt på det nominelle niveau). Projektet har desuden vist at BMD-metoden er fleksibel nok til at håndtere analyse-mæssige problemstillinger i økotoxikologien, også for andre typer af data end der hvor metoden allerede anvendes på i dag. Dette er illustreret gennem typiske data eksempler fra økotoxikologien.



Miljøstyrelsen  
Tolderundsvej 5  
5000 Odense C

[www.mst.dk](http://www.mst.dk)