

# Statistisk analyse og biologisk tolkning af toksicitetsdata

Jens C. Streibig  
Institut for Jordbrug og Økologi

Christian Ritz, Torben Martinussen & Nina Cedergreen  
Institut for Grundvidenskab og Miljø

Anne Munch Christensen  
Institut for Farmaci og Analytisk Kemi

Anders Baun & K. Ole Kusk  
Institut for Vand og Miljøteknologi  
Danmarks Tekniske Universitet

Christian Friis  
Institut for Veterinær Patobiologi  
Københavns Universitet

Miljøstyrelsen vil, når lejligheden gives, offentliggøre rapporter og indlæg vedrørende forsknings- og udviklingsprojekter inden for miljøsektoren, finansieret af Miljøstyrelsens undersøgelsesbevilling.

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.

# Indhold

FORORD	5
SAMMENFATNING	7
SUMMARY	9
1 INTRODUKTION	11
1.1 PROBLEMSTILLING OG AFGRÆNSNING	14
2 DOSIS-RESPNSMODELLER	17
3 PROGRAM TIL STATISTISK DATABEHANDLING	25
4 PROBLEMER, DER ER UAFHÆNGIGE AF TYPEN AF RESPONS	27
4.1 SAMMENLIGNING AF ED/EC-VÆRDIER	27
4.1.1 <i>Problemstilling</i>	27
4.1.2 <i>Datakilde</i>	27
4.1.3 <i>Analyse af data</i>	27
4.1.4 <i>Sammenfattende bemærkninger</i>	29
4.2 DESIGN AF BIOASSAYS: FLERE DOSERINGER KONTRA FLERE REPLIKATER	30
4.2.1 <i>Problemstilling</i>	30
4.2.2 <i>Datakilde</i>	30
4.2.3 <i>Simulationsstudie</i>	30
4.2.4 <i>Sammenfattende bemærkninger</i>	34
4.3 MODELVALGETS INDFLYDELSE PÅ ESTIMATER FOR EC-VÆRDIER	35
4.3.1 <i>Problemstilling</i>	35
4.3.2 <i>datakilde</i>	35
4.3.3 <i>Simulationsstudie</i>	35
4.3.4 <i>Sammenfattende bemærkninger</i>	37
5 ANALYSE AF KONTINUERT RESPONS	39
5.1 VARIANSHETEROGENITET	39
5.1.1 <i>Problemstilling</i>	39
5.1.2 <i>Datakilde</i>	39
5.1.3 <i>Sammenfattende bemærkninger</i>	42
5.2 JUSTERING FOR VARIATION FRA KURVE TIL KURVE	42
5.2.1 <i>Problemstilling</i>	42
5.2.2 <i>Datakilde</i>	42
5.2.3 <i>Sammenfattende bemærkninger</i>	44
6 ANALYSE AF BINOMIALT RESPONS	45
6.1 PROBLEMSTILLING	45
6.1.1 <i>Datakilde</i>	45
6.1.2 <i>Analyse af data</i>	45
6.1.3 <i>Afsluttende bemærkninger</i>	48

6.2	ANALYSE AF BINOMIALFORDELTE RESPONSMODELLER FOR KONTINUERT RESPONS	49
6.2.1	<i>Problemstilling</i>	49
6.2.2	<i>Datakilde</i>	49
6.2.3	<i>Simulationstudie</i>	49
6.2.4	<i>Afsluttende bemærkninger</i>	50
7	"LACK OF FIT"-TEST	51
7.1	TEST BASERET PÅ KUMULEREDE RESIDUALER	51
7.2	AFSLUTTENDE BEMÆRKNINGER	52
8	SEMI- OG IKKE-PARAMETRISK DOSIS-RESPONS-MODELLERING	53
8.1	ALTERNATIVE DOSIS-RESPONSMODELLER	53
8.2	AFSLUTTENDE BEMÆRKNINGER	54
9	EKSEMPLER PÅ DOSIS-RESPONS PROBLEMSTILLINGER	55
9.1	ØVRE OG NEDRE ASYMPTOTER	55
9.1.1	<i>Problemstilling</i>	55
9.1.2	<i>Datakilde 1 (bioassay med blandet bakteriekulturer)</i>	55
9.1.3	<i>Datakilde 2 (pesticider og brystcancer celler)</i>	57
9.1.4	<i>Afsluttende bemærkninger</i>	58
9.2	FØRSKELLIGE ENDPOINTS	58
9.2.1	<i>Problemstilling</i>	58
9.2.2	<i>Datakilde 3</i>	59
9.2.3	<i>Datakilde 4</i>	60
9.2.4	<i>Afsluttende bemærkninger</i>	62
9.3	MODIFICEREDE RESPONSKURVER	62
9.3.1	<i>Datakilde 5</i>	64
9.3.2	<i>Afsluttende bemærkninger</i>	66
10	TO-GENERATIONSREPRODUKTIONSFORSØG	67
10.1	PROBLEMSTILLING	67
10.1.1	<i>Kuld størrelse</i>	67
10.1.2	<i>Afsluttende bemærkninger</i>	70
11	GENEREL DISKUSSION	71
12	KONKLUSION	75
13	PERSPEKTIVERING	77
	LITTERATUR	79
	BILAG A: REGNEEKSEMPEL MED DRC	83

# Forord

Projektgruppens kommissorium er at komme med retningslinjer for korrekt og tidssvarende statistiske analyser af toksicitets- og økotoksicitetsdata.

Projektdelegerne har store mængder data fra forsøg fra forskellige projekter inden for såvel pesticiders som andre miljøfremmede stoffers toksikologi og økotoksikologi. Projektets mål er generelt at give vejledning i anvendelsen af den optimale og korrekte anvendelse af statistiske metoder og statistiske modeller og herunder at kunne rådgive Miljøstyrelsens administration i, hvorledes forskellige forsøgsdata med forskellige statistiske karakteristika bør analyseres, så de parametre, der benyttes ved risikovurdering, afspejler den naturlige variation bedst muligt.

Efter første projekt år forlod Anne Marie Vinggaard, DTU, gruppen og relativt sent i forløbet fik vi suppleret gruppen med Christian Friis fra KU.

Projektet har været fulgt af en stygruppe med følgende sammensætning

Peter B. Sørensen, AU  
Aage Vølund, Novo Nordisk  
Mogens Erlandsen, AU  
Ulrik Nørum, SDU  
Claus Hansen, MST  
Susanne Hougaard, MST  
Jørn Kirkegaard, MST

Vi vil gerne takke styringsgruppens medlemmer for deres engagement i projektet og for alle de konstruktive forslag, de er fremkommet med gennem hele projektperioden.

Vi håber at rapporten kan blive brugt som inspiration for videreudvikling af statistiske analysemetoder af dosis-responsdata til måling af toksicitet. Det er samtidig vores håb, at rapporten kan bruges af studerende, forskere og andre, som arbejder med at planlægge, udføre og analysere dosis-responseeksperimenter inden for toksikologi og økotoksikologi.



# Sammenfatning

I dette projekt er det undersøgt hvordan man ved at anvende egnede statistiske metoder kan optimere anvendelsen af eksisterende eksperimentelle toksikologiske og økotoksikologiske data for pesticider. Projektet beskæftiger sig overvejende med sigmoide dosis-responsmodeller, såvel symmetriske som asymmetriske. Disse er anvendt på både på data med kontinuerte udfald, som for eksempel resultater af test med vækst af planter, mikroorganismer eller dyr, og test med binære udfald, som for eksempel antallet af døde eller immobile dyr. Dosis-responskurver er karakteriseret ved enten at være monotont stigende eller faldende fra en øvre/nedre asymptote til en nedre/øvre asymptote. Man er ofte interesseret i at bestemme den dosering eller koncentration, der medfører en virkning/effekt som ligger midt mellem doseringskurvernes øvre og nedre asymptote. Denne dosering eller koncentration benævnes EC50, ED50, LC50 eller LD50 afhængigt af om man anvender koncentrationer (EC50, LC50), doser (ED50, LD50) og om man er interesseret i letale (LC50, LD50) eller andre typer af effekter (EC50, ED50). Kvantificering af toksicitet i administrativ sammenhæng er ofte knyttet til EC50/LC50/ED50/LD50, og inden for økotoksikologien også til de doser/koncentrationer, der giver 10% effekt (EC10/LC10/ED10/LD10). Disse værdier er alle estimater, der baserer sig på en matematisk beskrivelse af dosis-respons forløbet samt en statistisk kvantificering af variansen på estimaterne hørende til kurvebeskrivelsen. Kvantificering af lave effekt-niveauer er også vigtige inden for humantoksikologien, men her benyttes ofte variansanalyse for at finde den såkaldte NOEL (No Observable Effect Level),. Herved udnytter man ikke hele dosis-respons forløbet, og en langt bedre dataudnyttelse kan opnås ved at forlade denne praksis og i stedet bruge regressionsanalyse. Det skal nævnes, at man af og til opnår resultater, der udviser ikke-monotone kurveforløb. Denne type resultater komplicerer definitionerne af estimater som EC50/LC50/ED50/LD50, og denne type data er behandlet særskilt i rapporten.

De udførte simuleringer af eksisterende dosis-respons kurver påviser, at præcisionen af toksicitetsparametrene bliver forbedret ved at sænke antallet af gentagelser indenfor doseringerne og i stedet udvide antallet af doseringer. Dette er særdeles relevant i praksis, da man ofte vil have begrænsninger i antallet af forsøgsheder på grund af plads eller økonomi. I det foreliggende datamateriale påviste vi desuden, at for EC50 fås stort set samme værdier og usikkerheder uanset om der benyttes en symmetrisk eller en asymmetrisk dosis-responskurve. Derimod kan valget af model få en betydning for estimation af effektkoncentrationer, der er placeret i yderenderne af kurven (fx EC10 og EC90).

Ved statistisk analyse af kontinuerte responsdata er det vigtigt at undersøge, om der er variansheterogenitet i datasættet. For de datasæt hvor dette var tilfældet, er der i denne rapport benyttet den såkaldte Box-Cox transformation af såvel venstre som højre side af regressionsmodellen med henblik på at sikre, at variansen på responsen er konstant uanset responsniveau. Selve parameterestimaterne, f.eks. EC50 og EC10, ændrer sig ikke synderligt, hvis man analyserer data uden at tage højde for variansheterogenitet. Derimod vil usikkerheden på EC50 og EC10 være påvirket, og usikkerheden på

parametrene bliver ikke bestemt korrekt. Det samme gør sig gældende, hvis man fejlagtigt analyserer binomialfordelte data, som om de er kontinuerte. Da ikke blot de nævnte parameterestimer, men også den tilhørende usikkerhed er af afgørende betydning for de vurderinger man foretager i forbindelse med pesticiders registrering, er det bydende nødvendigt, at data analyseres med de korrekte statistiske analyser..

I rapporten behandles også semi- og ikke-parametriske analyser, dvs. analyser hvor der ikke på forhånd er nogen klar forestilling om, hvorledes dosis-responsforløbet bør modelleres. Ligeledes indeholder rapporten en gennemgang af problemstillinger, der opstår med dosis-responskurvers forskellige nedre og øvre asymptoter samt problemstillinger vedrørende sammenligning af data fra forsøg med flere forskellige responsvariable. Som eksempel anvendes et to-generationers reproduktionsforsøg med rotter designet til at blive analyseret efter NOEL principper. Her viste en analyse med en log-logistisk model, at der var store forskelle mellem NOEL og en bench mark dosis baseret på dosis-responsmodellen.

Flere af de i dag anvendte statistiske metoder til vurderingen af toksikologiske og økotoksikologiske data i administrativ sammenhæng er ofte utidssvarende – og i visse tilfælde endda ukorrekte. For at ændre den gængse praksis kræves det at de regulerende myndigheder presser på i internationale fora. De statistiske analyser, der er gennemført i denne rapport, kan danne det videnskabelige grundlag for at foreslå ændringer af gammel, men velgennemprøvet praksis, således at det fulde potentiale af resultater af eksisterende toksikologiske og økotoksikologiske tests udnyttes. Dette kan f.eks. være ved at revidere fastlåste meninger om antallet af gentagelser i regressionsammenhæng og ved at sikre at datas beskaffenhed (binomialt eller normalt fordelte data) analyseres med korrekte statistiske metoder.



# Summary

This project is aimed at studying how the application of sigmoidal dose-response models (symmetrical and asymmetrical) combined with proper statistical analysis can be used to optimize the interpretation of existing experimental data obtained in toxicological and ecotoxicological tests of pesticides. These models are used to describe the dose-response relationships for continuous data, e.g. growth, biomass or enzyme concentration/inhibition, and binomially distributed data, e.g. dead or live animals and plants, mobile or immobile animals. Dose-response curves are either monotonically increasing from a lower to an upper asymptote or decreasing from an upper to a lower asymptote. Usually, we are interested in determining the dose or concentration of a pesticide that yields a 50% reduction in the response relative to the upper and lower asymptotes. This dose or concentration is termed EC50, ED50, LC50, or LD50 depending on whether it is the concentration (EC50, LC50) or dose (ED50, LD50) that is used and whether the endpoint is lethality (LC50, LD50) or other types of effects (EC50, ED50).

For regulatory uses the quantification of toxicity is often related to EC50/LC50/ED50/LD50, but in ecotoxicology also those concentrations/doses yielding low effects, e.g. 10% (EC10/LC10/ED10/LD10), are of interest. All these values are estimates based on a mathematical description of the dose-response relationship combined with a statistical quantification of the variance on parameter estimates. In human toxicology the quantification of doses/concentrations for which no or low effects occur is also of high importance. Here the use of analysis of variance usually the statistical method recommended for determining the so-called NOEL (No Observable Effect Level). However, by doing so the whole dose-response relationship is not used. By using proper regression a far better basis for interpretation of the experimental results will be obtained.

On the basis of simulations of existing dose-response curves we showed that if there are limits to the number of experimental units, either because of limitation in space, economy or test organisms, the precision of the parameter estimates for toxicity will be improved by reducing the number of replications at each dose and expanding the number of doses. In the available data material we showed that the EC50/LD50/EC50 is virtually the same whether one applies a symmetric or an asymmetric dose-response curve. However, the choice of regression model does play a role if the EC10 or EC90 response levels are of interest.

For continuous responses it is important to ensure that there is homogeneity of variance. If this prerequisite does not apply it is required to use an appropriate method to transform data so they comply with this prerequisite. In this report we have used the so-called Box-Cox transformation of both sides of the regression model (also called "Transform-both sides"). The parameter estimates of say EC50 or EC10 do not change dramatically whether we use data with variance heterogeneity or not, but the standard errors of the parameters would not be correct and sometimes will differ

considerably from the ones obtained from the analysis with a Box-Cox transformation that takes the heterogeneity of variance into account. The same problems arise when one analyses binomially distributed data as if they are coming from a continuous distribution. As the estimated EC50/ED50/LD50 or EC10/ED10/LD10 and their associated standard errors are of primary importance when registering a pesticide and place it in toxicity classes, it is imperative to apply proper statistical methods.

The report also touches upon semi and non-parametric analysis, i.e. analyses that do not have any *a priori* functional dose-response specification. Furthermore, the report also looks at various problems with definitions of upper and lower asymptotes and how to compare experiments with several measured responses. One of the examples in the report is a dose-response experiment based upon generation studies in rats were also included. The protocols were designed for a No Observable Effect Level (NOEL) study whilst we analysed them using dose-response regression analysis and obtained more useful information.

It is concluded that in order to exploit the full potential of well-established statistical methods, it is important that regulatory authorities take initiatives to replace obsolete and insufficient methods and concepts, and require that appropriate methods, which account for the statistical properties of the data, must be used.

# 1 Introduktion

For at dokumentere pesticiders biologiske virkning undersøger man deres virkning på levende organismer, levende væv eller lignende. De forsøg man laver kaldes biologiske assays eller bioassays, og deres formål er at måle et givet stofs stimulering eller reduktion af livsytringer i dyr og planter. Det kan være *in vivo* forsøg med dyr og planter og/eller *in vitro* forsøg med celler, enzymsystemer mv. Fælles for bioassays er, at de beskriver sammenhængen mellem virkning og dosis (Finney 1978; Hewlett and Plackett 1979). OECD, ISO og andre internationale organisationer og institutioner publicerer løbende dokumenter, som giver retningslinjer for det eksperimentelle design af bioassays, for hvordan toksicitet skal måles og for hvordan kemiske analyser af stoffer anvendt i testene skal udføres (fx ISO 6341 1989; ISO13829 1997; ISO 8692 2004; OECD 208; OECD 227). Derimod er retningslinjerne for de statistiske analyser til kvantificering toksicitet ofte mere uklare.

Godkendelse og brug af pesticider er reguleret af rammer for vurdering af plantebeskyttelsesmidler, som er forankret i EU-direktivet 91/414/EØF. Uanset om et stof er helt nyt på markedet eller om det indgår i et revurderingsprogram er der krav om en lang række humantoksikologiske og økotoksikologiske undersøgelser. Disse anvendes i en risikovurdering af det pågældende stof til at fastslå om stoffet udgør en risiko for sundhed og miljø. Registreringsmyndighederne benytter desuden resultaterne af disse undersøgelser til at fastslå, om et givet stof overholder givne afskæringsværdier for toksicitet og økotoksicitet. Således har resultaterne af bioassay afgørende betydning for godkendelse af pesticider, og da disse bioassays desuden ofte er meget omkostningstunge er det vigtigt at de opnåede resultater udnyttes fuldt ud og i overensstemmelse med god videnskabelig praksis. Grundlaget for nærværende projekt var derfor at kigge nærmere på den statistiske analyse af resultater opnået i gængse standardiserede testtyper. Målgruppen for denne rapport er alle som beskæftiger sig med at fremskaffe og vurdere toksicitets- og økotoksicitetsdata. Projektet gennemgår således flere af de væsentligste problemstillinger, man kommer ud for, når man skal designe, udføre, analysere og vurdere et bioassay.

I princippet er der tre forskellige måder at analysere bioassay-data på:

1. Variansanalyse: Har til formål at analysere forskelle mellem middelværdier af i forvejen fastlagte behandlinger. Her kan man for eksempel undersøge om der er forskelle mellem effekter af udvalgte doseringer set i forhold til effekter i en ubehandlet kontrolgruppe. Man har altså ikke på forhånd en forventning om en bestemt sammenhæng mellem virkning og dosis, og formålet er ikke at interpolere mellem doser. Man forudsætter blot, at der til hver dosis er tilknyttet en parameter, et niveau, som beskriver virkningen af netop den dosis. Man er interesseret i at finde den dosering i en stigende doseringsrække, der netop ikke giver en virkning på testorganismene. Kriteriet for, at der ikke er virkning er et *a priori* valgt signifikansniveau. Denne koncentration/dosis kaldes ofte NOEC/NOEL (No Observable Effect Concentration/Level). På tilsvarende vis fastlægges også LOEC (Lowest Observable Effect Concentration), der defineres som den dosering i en doseringsrække, der netop har en statistisk signifikant virkning på organismene. Brug af NOEC og LOEC forudsætter dermed, at der findes

en tærskeldosering som netop udløser en virkning på en organisme. Desuden vil NOEC og LOEC kun kunne antage værdier, der modsvarer en af de koncentrationer, der indgår i det eksperimentelle testdesign. Inden for toksikologien taler man ofte om niveauer i stedet for koncentrationer, så i stedet for NOEC, som anvendes i økotoksikologisk sammenhæng, anvender man NOEL (No Observable Effect Level) for human toksikologiske test.

2 Parametrisk dosis-respons model: Denne metode forudsætter, at man enten *a priori* eller *a posteriori* har en formodning om, at virkningen af et givet pesticid kan beskrives via en parametrisk funktionel sammenhæng, enten lineær eller ikke-lineær. Når man udfører en toksicitetstest af et pesticid i et passende stor doseringsområde vil data ofte have et ikke-lineært forløb, når responset bliver afbildet som funktion af dosis. Med undtagelse af specielle områder inden for toksikologi og økotoksikologi er der enten tale om monotont stigende eller faldende doseringskurver med asymptotiske, øvre og nedre grænser (Finney 1978). Med andre ord kan vi sige, at ved at bruge regressionsanalyse - ofte ikke-lineær - til at beskrive en dosis-responskurve, antager vi, at der altid er en virkning af et stof uanset hvor lille dosering man benytter. Altså strider definitionen af NOEC og LOEC (at der findes en dosis som ingen virkning har (NOEC) og en dosis som netop har en virkning (LOEC)) mod regressionsanalysens forudsætninger. Hvis man vil definere et nul-effekt niveau (ofte anvendt som en tilnærmelsesvis NOEC eller LOEC) ud fra dosis-responskurver, skal man derfor definere et vist responsniveau, som ud fra et videnskabeligt eller administrativt synspunkt anses for acceptabelt til at beskrive, at der ikke er en effekt.

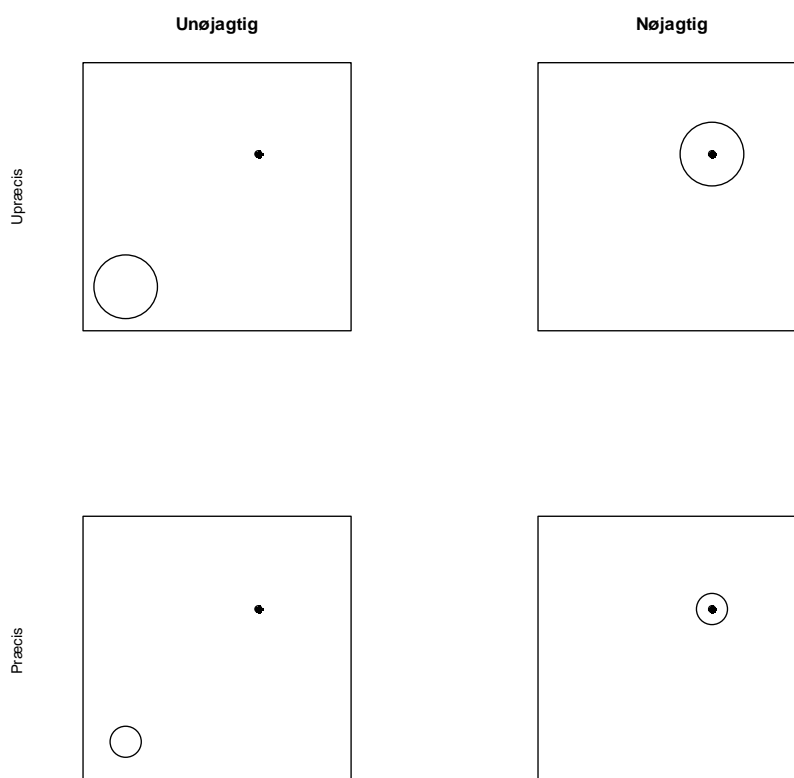
3 Semi-parametrisk model: Den tredje metode er at benytte semi-parametriske dosis-responsmodeller. Her har man ingen *a priori* opfattelse af givne parametriske sammenhænge eller kun vage forestillinger om sammenhæng mellem respons og dosis.

Variation i testresultater er en naturlig del af et bioassay. Den systematiske variation, der beskrives af dosis-respons sammenhænge, skyldes virkningen af et stof på en organisme, mens den resterende variation kan skyldes en lang række årsager, for eksempel genetisk variation mellem individer, opvækstvilkår for forsøgsorganismer, variation i forsøgsbetingelserne og sidst, men ikke mindst, målefejl og -unøjagtigheder. Ved hjælp af forsøgsplaner, der skal afbøde uvedkommende systematisk indvirkning på responsen, vil et veltilrettelagt dosis-responseksperiment tage hensyn til en lang række af de faktorer, der giver anledning til variation (Finney, 1978; van Leeuwen & Hermens, 1995). For at minimere den tilfældige variation har nationale og internationale standardiseringsorganisationer, (Dansk Standard, OECD, ISO m.fl.) givet retningslinier for datakrav, og som hovedregel er der krav om, at bioassay til registreringsformål udføres af autoriserede GLP laboratorier (Good Laboratory Practice).

Med højerestående dyr benytter man som regel kun få doseringer (f.eks. 3 doseringer) af såvel økonomiske som dyreetiske årsager. Yderligere har der været en stor indsats for at substituere dyreforsøg med alternative metoder (ECVAM <http://ecvam.jrc.it/index.htm> European Centre for Valuation of Alternative Methods) ved at benytte surrogatorganismer/dele af organismer, hvis reaktioner på pesticidpåvirkning er højt korreleret med en given dyregruppes påvirkning. For at styrke fortolkningen af resultater fra dyreforsøg, vil man ofte supplere disse med et batteri af cellebaserede og enzymbaserede laboratorietest, der målrettet kan undersøge for forskellige effekter og virkemekanismer. I denne forbindelse kan bl.a. nævnes gentoksiske og hormonlignende effekter (ISO 13829 1997; Soto *et al.* 1995). For planter

og mikroorganismer er det sjældent etiske overvejelser eller prisen på forsøgsindivider, der sætter en grænse for størrelsen af et assay, men derimod pladsmangel. F.eks. antallet af brønde i titerplader, plads i vækstkamre og lysforhold, etc.

Uanset hvad man måler er spørgsmålet om nøjagtighed og præcision afgørende for tolkningen af resultaterne. Da disse to begreber ofte sammenblandes eller ikke bruges konsekvent har vi i Figur 1.1 illustreret hvordan ordene "nøjagtighed" og "præcision" vil blive anvendt i denne rapport. Nøjagtighed er således stort set uafhængig af statistik, og drejer sig udelukkende om det er den "sande værdi" man måler. Det kan være en måleparameter i sig selv eller en værdi afledt af en måleparameter. Viden om den værdi man ønsker at måle er derfor afgørende nødvendig, og kan ikke erstattes af statistiske metoder. Præcision er derimod et mål for variationen af en given måling eller af en værdi afledt heraf. Det er netop præcisionen, der kan forbedres ved at lave et veltilrettelagt bioassay, hvor man tager højde for ydre og indre variationsårsager i selve forsøgsplanen. Figur 1.1 viser, at præcision og nøjagtighed ikke nødvendigvis hænger sammen. Det kan skyldes, at man ikke måler, det man tror man måler. Andre gange er det et bevidst valg, fordi man vælger en respons, der er nem at måle og som formodes at være højt korreleret med den respons, man anser for vigtig.



Figur 1.1 Sammenhængen mellem nøjagtighed og præcision. Det sorte punkt illustrerer den sande værdi, som kan være en parameter eller en afledning heraf. Størrelsen af cirklen illustrerer præcisionen som for eksempel kan være givet ved et konfidensinterval.

## 1.1 Problemstilling og afgrænsning

Dette projekt er målrettet mod at beskrive dosis-respons ved hjælp af parametriske ikke-lineære regressionsmodeller. Projektets formål er at komme med retningslinjer for korrekt og tidssvarende statistiske analyser af toksicitets- og økotoksicitetsdata under anvendelse af allerede eksisterende data fra forsøg udført af projektdeltagerne i forskellige projekter rettet mod kvantificering af både pesticiders og andre biologisk aktive stoffers human- og økotoksikologiske virkninger. Projektets vil hermed kunne give såvel toksikologer/økotoksikologer som administrative brugere af testresultater et indblik i de metoder, der er udviklet i de sidste år indenfor analyser af bioassay.

Rapporten er fokuseret på parametriske regressionsmodeller og de problemstillinger, som projektgruppens medlemmer i praksis har identificeret gennem deres arbejde med bioassays. En del af de problemstillinger, man ofte kommer ud for, når man analyserer bioassays, er nævnt i nedenstående tabel 1.1. Det skal dog understreges, at der ikke er tale om en udtømmende liste og at de nævnte problemstillinger ikke er isolerede, men ofte griber ind i hinanden.

Tabel 1.1. Identificerede problemstillinger ved statistisk analyse af resultater fra bioassays.

Problemer, der er uafhængige af typen af respons

*Sammenligning af ED/EC-værdier*

*Design af bioassay: Flere doseringer kontra flere replikater*

*Modelvalgets indflydelse på estimater for EC-værdier*

Analyse af kontinuert respons

*Variansheterogenitet*

*Justering for variation fra kurve til kurve*

Analyse af binomialt respons

Analyse af binomialfordelt respons med modeller for kontinuert respons

Tests for "Lack of fit"

Semi- og ikke-parametrisk dosis-respons-modellering

Forskellige eksempler

*Øvre og nedre asymptoter*

*Forskellige endpoints*

*Modificerede respons kurver*

To-generationsreproduktionsforsøg

Nogle de problemstillinger, som projektet indledningsvist havde identificeret er ikke eksplicit blevet behandlet i nærværende rapport: NOEC/LOEC-problematikken er således ikke behandlet yderligere, men det skal dog nævnes, at netop anvendelsen af variansanalyse og afhængigheden af et godt testdesign, har medført en så udtalt kritik af brugen af NOEC/NOEL og tilsvarende LOEC/LOEL til fortolkning af såvel toksikologiske som økotoksikologiske resultater (f.eks. Laskowski, 1995 og Kooijman, 1996), at vi finder disse parametre uanvendelige set fra et videnskabeligt synspunkt. Det har i løbet af projektet ikke været muligt at fremskaffe data, som var anvendelige til at illustrere variabilitet dosis, som er den uafhængige variabel, selvom denne variabilitet er særdeles vigtig ved analyse af bioassays. Ligeledes har vi i projektet afgrænset os fra at inkludere analyser af tidsserier, da der i projektgruppen ikke var rådighed over data som på passende vis kunne anvendes til at adressere dette problem. Med hensyn til problematikken om øvre og nedre grænse i bl.a. kombinationsforsøg vil vi henvise til (Ritz *et al*, 2005).

De problemstillinger, der bliver behandlet i denne rapport vil således afspejle gruppens arbejde med bioassays inden for forskellige discipliner. I rapporten har vi valgt først at se på de mere generelle og teoretiske statistiske aspekter af databehandling og fortolkning af bioassays for derefter at give praktiske eksempler, der skal tjene til at illustrere de teoretiske overvejelser.

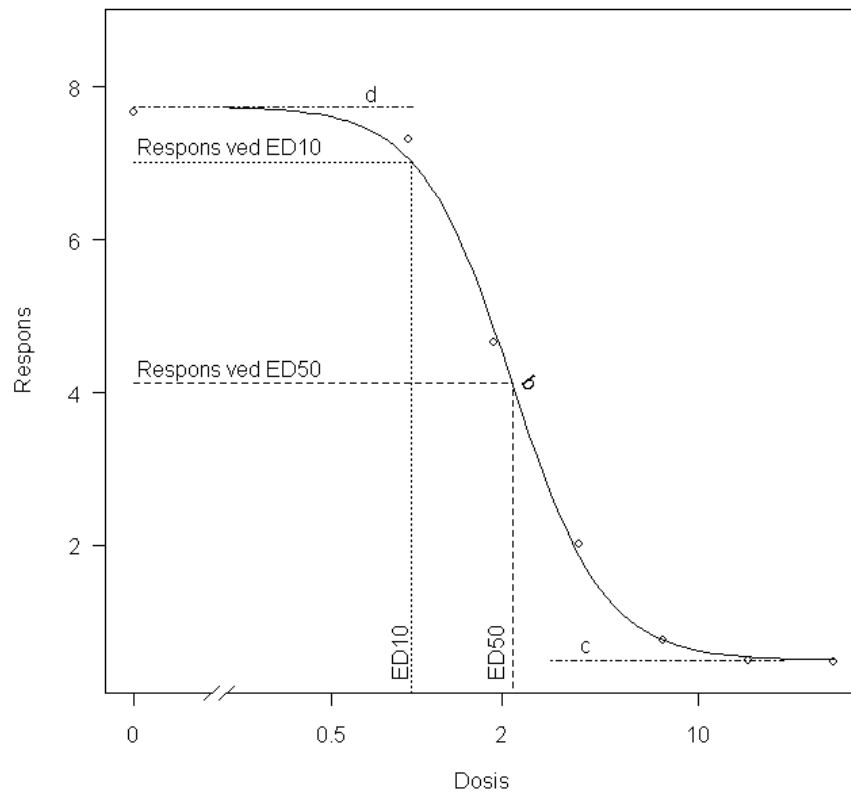




## 2 Dosis-responsmodeller

Der er en lang række parametriske modeller tilgængelige til at beskrive dosis-responsdata. I dette afsnit introduceres de mest almindeligt anvendte parametriske dosis-responsmodeller inden for toksikologien. Disse modeller har ofte et s-formet (sigmoidt) kurveforløb, men det skal nævnes at der også findes andre modeller så som log-dosis lineære, eksponentielle og ikke-monotone dosis-responsmodeller.

Det er fælles for mange af de inkluderede dosis-responsmodeller, at modelparametrene har en biologisk fortolkning. I Figur 2.1 kan man se en grafisk fremstilling af en karakteristisk dosis-respons kurve, hvor responsen nærmer sig en øvre grænse, når doseringen nærmer sig 0 (ubehandlet kontrol, dvs. dosis 0), og en nedre grænse, når doseringen bliver stor (f.eks. gående mod en fuldstændig hæmning ved høje doseringer). Den øvre og nedre grænse betegnes ofte som asymptoter. Den dosering eller koncentration, der fremkalder en virkning midt imellem de to asymptoter, er den dosering/koncentration, som giver en effekt på 50% i forhold til de øvre og nedre grænser. Denne dosering betegnes ofte ED50/EC50, hvor "E" refererer til en målt effekt og "D" hhv. "C" refererer til dosis hhv. koncentration. Figur 2.1 er et eksempel på data opnået i en test med kontinuert respons (f.eks. vækstratehæmning), men tilsvarende overvejelser gælder for diskret respons (f.eks. test med responset død/levende), hvor pedanten til ED50/EC50 ofte betegnes LD50/LC50 ("L" står for lethal altså dødelig). Til kvantificering og klassificering af et kemikalies effekt opsummeres de toksikologiske data ofte vha. 50%'s værdierne (LC50/LD50/EC50/ED50), men inden for human- og økotoksikologien kan der også være andre responsniveauer man er interesseret i, f.eks. effekter i det lave område, hvor det ofte er 10%'s værdierne, der anvendes (LC10/LD10/EC10/ED10). Disse er også vist i Figur 2.1.



Figur 2.1. Illustration af en log-logistisk dosis-responskurve, hvor den øvre ( $d$ ) og den nedre ( $c$ ) grænse er vist. Hældningen på kurven svarende til doseringen ED50 er angivet ved parameteren  $b$  (se model 2.1).

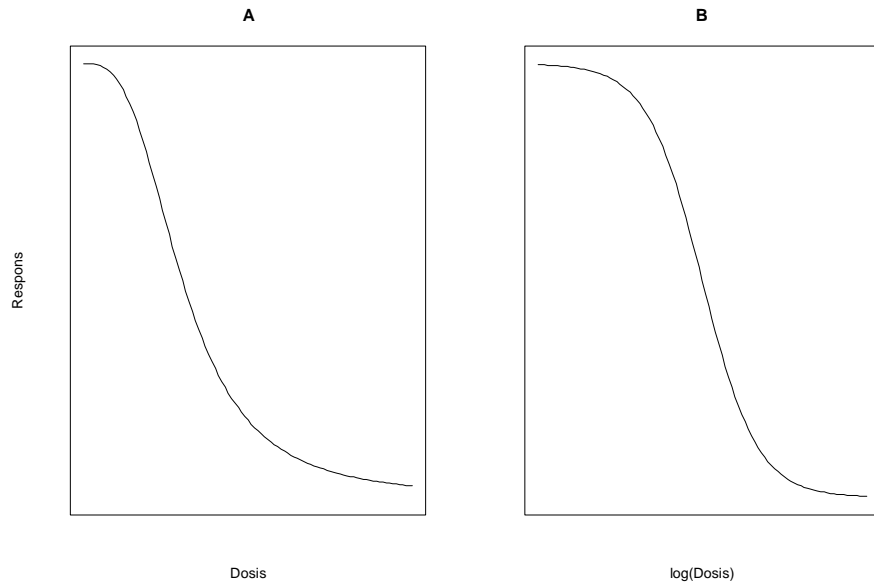
De sigmoide dosis-responsmodeller omfatter to typer, som ofte anvendes til at beskrive henholdsvis symmetriske og asymmetriske kurveforløb for monotont faldende eller stigende dosis-responsdata:

- 1) Den log-logistiske model vist i Figur 2.2 og ligning 2.1
- 2) De to Weibull-modeller vist Figur 2.3 og ligning 2.2. og 2.3

Som et særligt vigtigt eksempel skal den log-logistiske fire-parameter model nævnes. Den er defineret ved ligningen

$$y = c + \frac{d - c}{1 + \left( \frac{x}{ED50} \right)^b} \quad (2.1)$$

De fire parametre i Ligning (2.1) betegnes henholdsvis  $b$ ,  $c$ ,  $d$ , og  $ED50$ . Parameteren  $ED50$  betegner, som tidligere nævnt, den dosering, der resulterer i en respons midt mellem den nedre grænse  $c$  og den øvre grænse  $d$ . Parameteren  $b$  er proportional med hældningen på kurven omkring  $ED50$  (Streibig *et al*, 1993). Parametrene er vist på Figur 2.1. Ligning (2.1) er monotont faldende, hvis parameteren  $b$  er positiv, og stigende, hvis  $b$  er negativ.



Figur 2.2. Den log-logistiske dosis-responskurve som funktion af dosis (A) og logaritme-transformeret dosis (B).

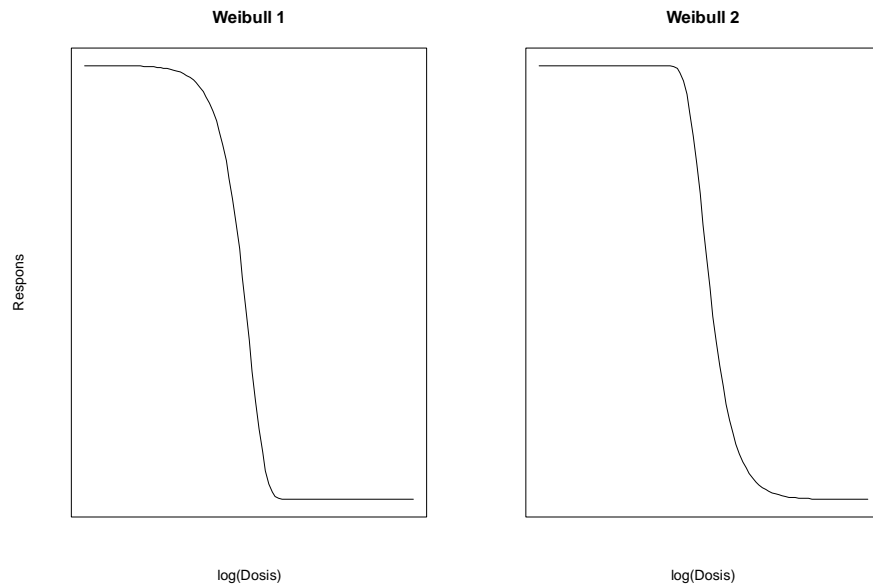
Som det fremgår af Figur 2.2 er kurven asymmetrisk, når man benytter en dosis-akse, der ikke er logaritmeret. Ligning (2.1) kan omskrives til Ligning (2.1a), som benytter  **$\log(x)$**  i stedet for  **$x$**

$$y = c + \frac{d - c}{1 + \exp(b(\log(x) - \log(ED50)))} \quad (2.1a)$$

Matematisk er Ligning (2.1) og (2.1a) identiske, men det er oftest formuleringen i ligning (2.1a), der benyttes i publikationer. Det er ofte muligt at antage, at den nedre grænse er 0, og den tilsvarende log-logistiske tre-parametermodel fås ved at sætte  $c=0$  i Ligning (2.1) og (2.1a). Den resulterende ligning er vist i Ligning (2.1b).

$$y = \frac{d}{1 + \exp(b(\log(x) - \log(ED50)))} \quad (2.1b)$$

De to Weibull-modeller er den anden hovedtype af modeller, der betragtes i dette projekt (Seber and Wild 1989, s. 338–339). De er kendetegnet ved at være asymmetriske, og Figur 2.3 illustrerer forskellene i kurveforløbene for de to Weibull-modeller. Den ene model har et kurveforløb, som er langsomt aftagende fra den øvre grænse (Figur 2.3 til venstre), mens den anden model har et forløb, som aftager hurtigt fra den øvre grænse (Figur 2.3 til højre).



Figur 2.3. De to Weibull-modeller er begge asymmetriske. Weibull-1-modellen er langsomt aftagende fra den øvre grænse, mens Weibull 2 er hurtigt aftagende fra den øvre grænse.

Begge Weibull-modeller har også fire parametre. Weibull-1-modellen er defineret ved ligningen

$$y = c + (d - c) \exp(1 - \exp(b(\log(x) - \log(e)))) \quad (2.2)$$

Weibull-2-modellen er givet ved ligningen

$$y = c + (d - c) \exp(-\exp(b(\log(x) - \log(e)))) \quad (2.3)$$

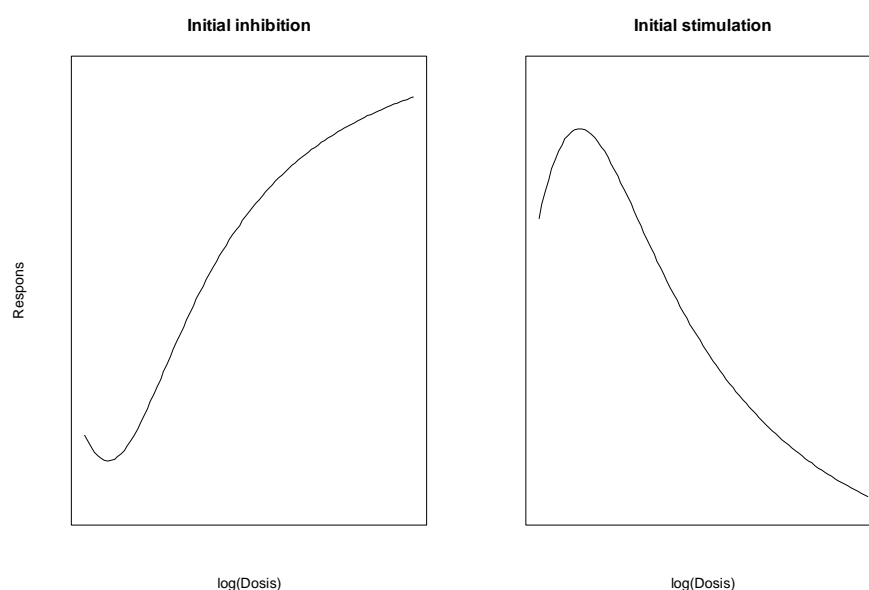
Parametrene  $c$  og  $d$  svarer til den nedre og øvre grænse ligesom for den log-logistiske model (Ligning (2.1)). Parameteren  $e$  svarer til den dosering, hvor dosis-responskurven har en vendetangent, men pga. det asymmetriske kurveforløb svarer denne dosering ikke til ED50. Parameteren  $b$  er proportional med hældningen på kurven omkring  $e$ . For begge Weibull-modeller vil en fastsat nedre grænse på nul ( $c=0$ ) reducere de fire parametre til tre parametre på samme måde som for den logistiske model (Ligning 2.1b). ED50 er som sagt ikke en parameter, der direkte indgår i Ligning (2.2) eller Ligning (2.3). Man kan dog stadig udregne ED50 på grundlag af parametrene i modellen.

I visse situationer er et observeret dosis-responsforløb ikke-monotont som funktion af dosis, f.eks. kan der ske en hæmning eller en stimulering af responsen ved meget små doseringer som vist i Figur 2.4. Dette benævnes hormesis (Cedergreen *et al.*, 2005). Denne type dosis-responsforløb kan beskrives ved en model afledt af den log-logistiske model i Ligning (2.1a) med følgende ligning

$$y = c + \frac{(d - c) \cdot f \cdot \exp(-1/x)}{1 + \exp(b(\log(x) - \log(e)))} \quad (2.4)$$

Ligning (2.4) giver anledning til en model med fem parametre, hvoraf kun  $d$  og  $c$  har samme fortolkning som i den log-logistiske model. Modellen kan

benyttes til at beskrive data, der udviser hormesis. Parametren  $f$  i Ligning (2.4) beskriver den maksimale størrelse på den hormesis effekt (Cedergreen *et al*, 2005).



Figur 2.4. Hæmning eller stimulering ved lave doseringer, som kan beskrives med forskellige modeller, der er afledt af den log-logistiske model i Ligning (2.1a). Ved initiale hæmning er parameteren  $b$  i Ligning (2.4) og (2.5) negativ, mens parameteren  $b$  for en initial stimulering er positiv.

I visse tilfælde kan en yderligere modifikation af Ligning (2.4) give en bedre beskrivelse af data med hormesis. Den modificerede model har følgende ligning

$$y = c + \frac{(d - c) + f \cdot \exp(-1/x^\alpha)}{1 + \exp(b(\log(x) - \log(e)))} \quad (2.5)$$

Ligning (2.4) svarer til at sætte  $\alpha=1$  i Ligning (2.5). Modellen baseret på Ligning (2.5) involverer seks parametre, hvor den sjette parameter er  $\alpha$  (Cedergreen *et al*, 2005). Hvis data ikke tillader en bestemmelse af  $\alpha$ , kan den fastsættes *a priori*, f.eks. som i Ligning (2.4). Andre er *a priori* valg af  $\alpha$  til at være 0,25 eller 0,5. Hvis der derimod er tegn på hormesis, det endvidere er muligt at tilpasse modellen baseret på Ligning (2.5), så kan man efterfølgende teste hypotesen  $\alpha = 1$  vha. et formelt statistisk test. Det er igen muligt at udregne ED50 ud fra modelparametrene under den antagelse, at niveauet til dosis 0 tages som den øvre grænse, hvilket i dette tilfælde ikke svarer til det størst mulige responsniveau. Praktiske erfaringer viser, at den maksimale respons i datasæt med hormesis ofte viser større variation end respons til dosis 0 (Cedergreen *et al*, 2005).

I Tabel 2.1 er der en forklaring på de parametre, som indgår i de ovennævnte ligninger.

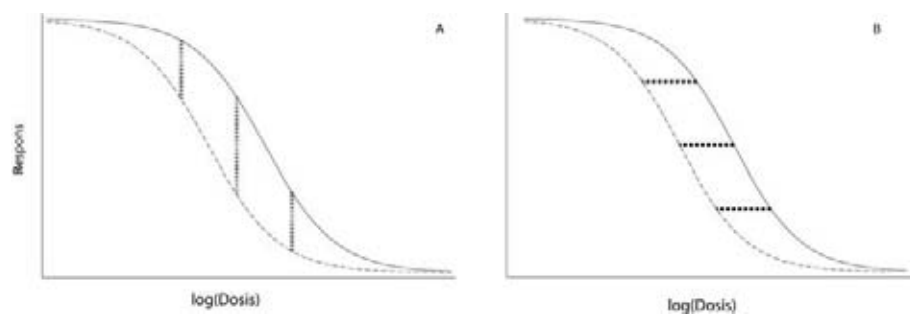
Tabel 2.1. Oversigt over biologisk fortolkning af parametrene i Ligning (2.1)-(2.5). Hormesis hentyder i denne sammenhæng til de ikke-monotone modeller i Ligning (2.4) og (2.5).

MATEMATISKE UDTRYK	MODELLER			
	LOG-LOGISTISK	WEIBULL 1	WEIBULL 2	HORMESIS
ØVRE GRÆNSE	D	D	D	D
NEDRE GRÆNSE	C	C	C	C
EC50	ED50	UDLEDES FRA LIGNING	UDLEDES AF LIGNING	UDLEDES AF LIGNING
MONOTON	JA	JA	JA	NEJ
SYMMETRISK	JA, PÅ LOG-DOSIS AKSE	NEJ	NEJ	NEJ

Ved at benytte de parametriske modeller, som vi har beskrevet ovenfor, opnår man den fordel, at man kan teste forskellige biologiske hypoteser. For eksempel kan man vha. statistiske metoder teste om man kan tillade sig at antage identiske parametre for dosis-responskurver. Imidlertid er forudsætningen for at benytte disse modeller, at der findes en funktionel sammenhæng. Fra et teoretisk synspunkt passer et sigmoidt kurveforløb ind i vores forestillinger om, hvorledes f.eks. en receptor vil reagere på en stigende dosis/koncentration af et kemikalie. Imidlertid vil der i mindre kontrollerede miljøer være tilfælde, hvor der ikke er så klare funktionelle sammenhænge mellem dosis og respons. I de tilfælde kan man forsøge sig med at benytte ikke-parametriske eller semi-parametriske metoder, som helt eller delvist giver slip på en underliggende parametriske model for det overordnede kurveforløb. Dette emne berøres kort senere i rapporten.

### ***Sammenligning af dosis-responskurver***

Som vist i Figur 2.5 er der i princippet to forskellige måder, hvorpå man kan sammenligne to dosis-responskurver: en vertikal sammenligning (som vist i Figur 2.5A) og en horisontal sammenligning (som vist i Figur 2.5B) (Streibig, 1988). Dosis-responskurverne kunne f.eks. vise to pesticiders virkning på en testorganisme, afprøvet under samme forsøgsbetingelser.



Figur 2.5. Den vertikale (A) og horisontale (B) sammenligning af to dosis-responskurver.

Den vertikale vurdering af to dosisresponskurver svarer i store træk til den måde man sammenligner effekter ved hjælp af en variansanalysemodel. I Figur 2.5A kan man se, at forskellene i effekter er særdeles påvirket af, hvilket dosisniveau man betragter. Det er klart ud fra Figur 2.5A, at forskellen mellem de to kurver (stiplede linjer) vokser som større doseringer, indtil

forskellene igen forsvinder ved meget høje doseringer, fordi responsen på et tidspunkt vil gå mod en fælles nedre grænse.

Figur 2.5B viser den horisontale vurdering. Forskellen mellem de to kurver ved et givet responsniveau er vist vha. stiplede linjer og svarer til den relative styrke mellem to stoffer. Som det fremgår af de to kurver i dette specielle tilfælde, Den horisontale forskel eller afstand er ens for de tre viste responsniveauer i Figur 2.5B, men sådan behøver det ikke altid at forholde sig. To kurver siges at være parallelle, hvis de har de samme modelparametre undtagen den, der bestemmer deres placering på dosisaksen.

Parallelforskydningen (afstanden mellem dosis-responskurverne) på  $\log(\text{dosis})$  akse vil være ens for hele kurveforløbet. Dette er den gængse måde, hvorpå man vurderer og sammenligner effekten eller styrken af forskellige kemikalier inden for toksikologi. I dosis-responsforsøg med mange dosis-responskurver, er der yderligere den fordel ved den horisontale vurdering, at man ofte kan reducere opsummeringen vha. få parametre, der beskriver, hvordan kurverne ligger forskudt relativt til hinanden. Det er ved at benytte statistiske metoder, at man kan reducere antallet af parametre til et minimum. Hvis data tillader, at man kan antage parallelle dosis-responskurver, så er det nok at give ét relativt styrkemål, uanset hvilket responsniveau der betragtes (Ritz *et al.*, 2006). Dette gælder også for responsniveauer tæt på den øvre og nedre asymptote. Eksempelvis er de to dosis-responskurver i Figur 2.5 parallelle, og derfor kan forskellen mellem dem opsummeres i et tal. Hvis kurverne ikke er parallelle, afhænger forskellene af hvilket responsniveau man betragter.





## 3 Program til statistisk databehandling

Pakken **drc** er en udvidelsespakke til statistikprogrammet **R**, som er open source og platform-uafhængig (<http://www.r-project.org>). Opbygningen af **R** gør det relativt nemt at lave udvidelsespakker, der specialiserer sig på områder, som ikke er dækket af den funktionalitet, som standardinstallation af **R** kommer med. Der findes lige nu mere end 1000 sådanne udvidelsespakker. Alle disse er underkastet ganske strenge krav vedrørende format, dokumentation og struktur, mens en egentlig validering af hver enkelt pakke typisk foregår gennem software-evaluering i specialiserede tidsskrifter og især gennem feedback fra brugere. Pakken **drc** har været igennem et software-evaluering i 2005 i forbindelse med publikationen (Ritz and Streibig, 2005). Funktionaliteten af **drc** er siden blevet væsentlig udvidet. På det seneste er det i høj grad sket gennem nærværende projekt, hvor problemstillinger og tilbagemeldinger fra projektgruppens medlemmer har bidraget til, at **drc** i dag fremstår som et væsentligt alternativ til eksisterende kommercielle programmer.

Pakken fokuserer på analyse af dosis-responsdata vha. regressionsmodeller for binær og kontinuert respons. I programpakken kan en række parametriske dosis-responsmodeller specificeres og estimeres, uden at brugeren skal bekymre sig om de tekniske aspekter, der ligger bagved (beregningsprocedurer, vælg af startværdier for parametrene i den ikke-lineære regression mm.). Estimationen er baseret på maximum likelihood-princippet. Dertil kommer funktionalitet, som ud fra en tilpasning af en dosis-responsmodel kan beregne vilkårlige ED-værdier eller relative styrker (**relative potencies**), når flere kurver betragtes. De tilhørende estimerede standardafvigelser og/eller konfidensintervaller kan også beregnes. Desuden er det muligt at visualisere data og modeltilpasningen grafisk. De fleste af de data der er vist i denne rapport findes som datasæt i **drc** pakken



## 4 Problemer, der er uafhængige af typen af respons

I dette afsnit vil vi gennemgå de problemstillinger vi har identificeret ved at analysere en lang række datasæt fra projektdeltagerne. Ikke alle problemstillinger er lige godt belyst med data, men vi har bestræbt os på at vælge bioassay fra de forskellige discipliner projektgruppen repræsenterer. Generelt kan det anføres at problemstillingerne alle er så generelle, at de ikke er disciplinorienteret. De statistiske problemer som toksikologen tumler med er stort set de samme som økotoksikologens. Forsøgsorganismer kan være forskellige, men problemstillingerne er ofte de samme.

### 4.1 Sammenligning af ED/EC-værdier

#### 4.1.1 Problemstilling

En af de oftest benyttede metoder til at vurdere pesticiders toksicitet og økotoksicitet er at sammenligne EC/ED-værdier fra forskellige, uafhængige bioassay/eksperimenter. Sådanne bioassay har ofte forskellige variansstrukturer, og så er spørgsmålet hvordan man bedst sammenligner?

#### 4.1.2 Datakilde

Som eksempel benyttes et dosis-respons-forsøg, der belyser effekten af de to herbicider bentazon og glyfosat ved en række doseringer. Formålet med forsøget var at undersøge de to herbiciders virkning på hvid sennep (*Sinapis alba*). Responsen er tørvægt i g/potte (Christensen *et al*, 2003). Bioassays for de to herbicider blev udført uafhængigt af hinanden. Data består af målte responser ved forskellige doser af hvert herbicid svarende til to dosis-responskurver. En 4-parameter log-logistisk model (Ligning (2.1a)) blev tilpasset data. Estimer for EC<sub>20</sub> på logaritmisk skala for bentazon og glyfosat var henholdsvis 3,11 (0,11) og 3,62 (0,15) med standardfejlen angivet i parentes. De tilbage-transformerede værdier er henholdsvis 22,3 med 95%-konfidensinterval (18,1 ; 27,6) og 37,3 med 95%-konfidensinterval (27,3 ; 50,8).

#### 4.1.3 Analyse af data

På grundlag af regressionsanalysen vil vi undersøge om de to EC<sub>20</sub> værdier er forskellige fra hinanden. Med andre ord, vi vil undersøge, om det kan antages, at differensen mellem estimerne for log(EC<sub>20</sub>) for de to herbicider er signifikant forskellig fra nul.

Standardfejlen på forskellen på log(EC<sub>20<sub>1</sub></sub>)-log(EC<sub>20<sub>2</sub></sub>) kan beregnes ud fra standardfejlen på de to estimer (som vi nedenfor henholdsvis kalder  $se_1$  og  $se_2$ ) vha. følgende formel:

$$\text{standardfejl p\aa differensen} = \sqrt{(se_1)^2 + (se_2)^2}$$

Effekten af forskellige variansstrukturer i to uafh\angige bioassay vil v\are inkluderet i de to standardfejl. Ud fra standardfejlen for differensen er vi i stand til at udregne et approksimativt 95%-konfidensinterval for differensen mellem de to log(EC20)-estimer:

$$\text{differens p\aa log(EC20)-estimer} \pm 1,96 \sqrt{(se_1)^2 + (se_2)^2}$$

I ovenst\aaende formel benyttes 1,96, dvs. 97,5%-fraktilen fra en standard normalfordeling. For kontinuerte data skulle man strengt taget i stedet for 1,96 bruge 97,5%-fraktilen fra en passende *t*-fordeling, men i langt de fleste tilf\ælde vil det ikke g\ore nogen v\asentlig forskel.

Baseret p\aa 95%-konfidensintervallet kan man s\aa teste (p\aa signifikansniveau 5%), om der er forskel p\aa de to EC-v\ardier ved at se, om 0 ikke er indeholdt eller er indeholdt i intervallet. Vi vil kalde denne test-procedure for differens-testet eller ratio-testet (Wheeler *et al*, 2006).

**Regneeksempel:** For ovenst\aaende data f\aa s f\olgende estimerede log(EC20)-v\ardier med standardfejl i parentes:

Bentazon: 3,11 (0,11)

Glyfosat: 3,62 (0,15)

Differensen p\aa estimerne er  $3,11 - 3,62 = -0,51$ , og den tilh\orende standardfejl vil v\are:  $\sqrt{(0,11)^2 + (0,15)^2} = 0,19$ . Et approksimativt 95%-konfidensinterval for differensen bliver s\aa:

$$-0,51 \pm 1,96 \cdot 0,19 = [-0,88 ; -0,14]$$

Konfidensintervallet indeholder ikke tallet 0, s\aa vi kan konkludere, at der er forskel p\aa EC20-v\ardierne for de to herbicider (p\aa et 5%-signifikansniveau).

Hvis man sammen med de estimerede EC-v\ardier f\aar oplyst de tilh\orende 95%-konfidensintervaller i stedet for standardfejl, kunne man v\are fristet til at sammenligne konfidensintervallerne direkte og konkludere, at der ikke er forskel, s\aa fremt de to intervaller overlapper. Wheeler *et al* (2006) frar\ader denne metode, fordi den har betydelig mindre styrke mht. at finde forskelle end differens-testet, der direkte sammenligner de to estimerede EC-v\ardier: Overlap-proceduren kan godt vise, at de to EC-v\ardier er ens, hvor de i virkeligheden (vurderet ud fra differens-testet som er vist ovenfor) er forskellige. Omvendt vil det v\are s\aa dan, at hvis man ud fra overlap-proceduren finder, at de to EC-v\ardier er forskellige, s\aa vil differens-testet ogs\aa vise, at der er forskel p\aa de to EC-v\ardier. Opsummerende m\aa det siges, at overlap-proceduren kan give misvisende resultater og derfor frar\ades det at bruge den.

**Regneeksempel fortsat:** Ved at tilbagetransformere fra logaritmisk skala fås følgende EC20-estimerer med tilhørende, individuelle konfidensintervaller:<sup>1</sup>

bentazon: 22,3 (18,1 ; 27,6)  
glyfosat: 37,3 (27,3 ; 50,8)

De to intervaller overlapper hinanden, og baseret på overlap-proceduren ville vi konkludere, at de to EC20-værdier ikke er forskellige fra hinanden. Men det er i modstrid med resultatet fra differens-testet. Altså giver overlap-proceduren anledning til et misvisende resultat.

Man kan ofte ud fra konfidensintervallet regne tilbage til den benyttede standardfejl og på den måde få fat i standardfejlene for de 2 estimerede EC-værdier. Fremgangsmåden er som følger:

1. Kontroller om intervallet er symmetrisk omkring den estimerede EC-værdi i lineær eller logaritmisk skala. Det kan nemlig kun lade sig gøre med symmetriske konfidensintervaller.
2. Tag det venstre endepunkt i intervallet og fratæk den estimerede EC-værdi eller log(EC)-værdi og divider derefter med -1,96 for binomialfordelte data (eller eventuelt en passende 97,5%-fraktil fra en  $t$ -fordeling for kontinuerte data). Det fremkomne tal er standardfejlen på den estimerede EC-værdi eller log(EC)-værdi.
3. Gentag det samme for det andet interval.
4. Herefter kan ovenstående procedure benyttes til at sammenligne EC- eller log(EC)-værdierne.

**Regneeksempel:** Antag er EC50 i et givet dosisrespons assay er estimeret til 16,6 med et 95%-konfidensinterval på [15,2 ; 18,0]. Så giver følgende udregning den tilhørende standardfejl:

$$(15,2 - 16,6) / (-1,96) = 0,71$$

#### 4.1.4 Sammenfattende bemærkninger

Såfremt både estimatet og den tilhørende standardfejl er angivet for dosis-responskurverne, så er det muligt at teste om de to EC-værdier kan være ens. Bemærk, at det er antaget, at de to forsøg, som skal sammenlignes, er uafhængige af hinanden. Så metoden kan ikke benyttes til at sammenligne forskellige endpoints fra samme test.

**Anbefaling:** Såfremt det er muligt at beregne et konfidensinterval for differensen mellem to logaritme-transformerede EC-værdier, så bør denne testprocedure benyttes. Kun i tilfælde hvor det ikke er muligt, bør overlap-proceduren benyttes.

---

<sup>1</sup> (  $\exp(3,11 - 1,96 * 0,11)$  ,  $\exp(3,11 + 1,96 * 0,11)$  ) for konfidensintervallerne og  $\exp(3,1)$  for parameterestimatet

## 4.2 Design af bioassays: Flere doseringer kontra flere replikater

### 4.2.1 Problemstilling

Et ofte diskuteret spørgsmål er, om det er bedre at have flere doseringer og færre replikater per dosis end omvendt? Vi betragter den situation, hvor der *a priori* er valgt et doseringsskema, hvorfra doseringer kan vælges. I praksis vil man ofte udføre et indledende studie med en koncentrationsrække, der spænder over flere dekader. Resultaterne herfra vil så blive brugt til at vælge doseringsrækken, som helst skal dække dosis der næsten ikke giver nogen virkning til doseringer der giver maksimal virkning. Hvis man overhovedet ikke har nogen idé om, hvilke doseringer der skal benyttes, så kan man f.eks. benytte metoden foreslået af Wright & Bailer (2007), som går ud på at man ud fra et valgt optimalitetskriterium bruger en computer-intensiv søgealgoritme til at finde de sæt af doseringer der bedst opfylder optimalitetskriteriet.

### 4.2.2 Datakilde

Vi vil benytte data fra et eksperiment, som undersøger virkningen af forskellige fenoler på rajgræs (*Lolium perenne L.*). Respons er rodlængde for varierende koncentrationer af ferulsyre i mM, dvs. vi betragter et eksempel med kontinuert respons. Data bruges til at estimere de parameterværdier som alle datasæt (i scenario A nedenfor) simuleres ud fra.

### 4.2.3 Simulationsstudie

Med en fire parameter log-logistisk regression fik vi en EC50-værdi på 3,06 mM. Ved forsøget blev følgende doseringer anvendt:

0,23; 0,47; 0,94; 1,88; 3,75; 7,50; 15; 30; 60 mM

Data bruges til at generere de parameterværdier, der er nødvendige for at foretage de simuleringer der vises nedenfor. Der blev genereret 1000 datasæt baseret på en log-logistisk model for hvert af de to scenarier, A og B er vist i Tabel 4.1.

Tabel 4.1. EC-værdier anvendt til datagenerering i de to scenarier til illustration af betydning af replikater i dosis-responsforsøg.

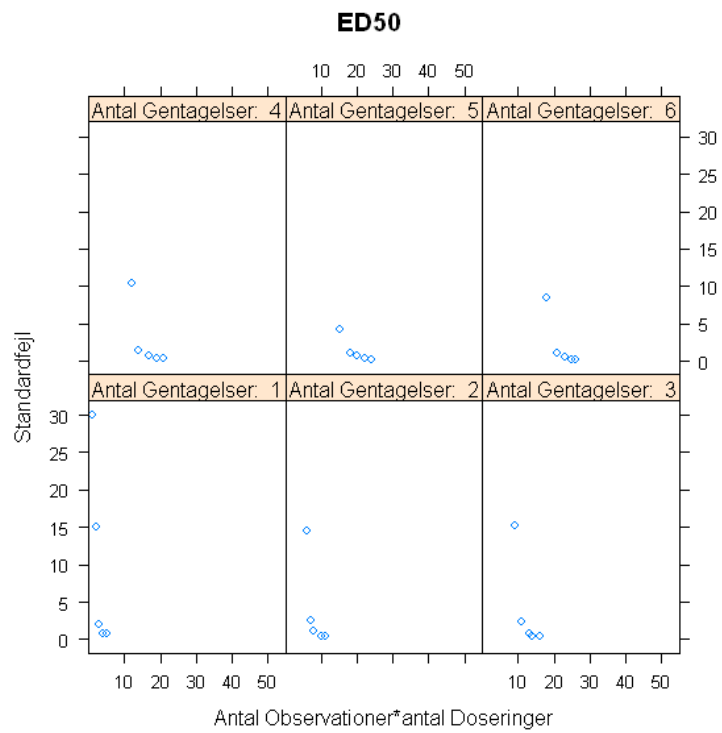
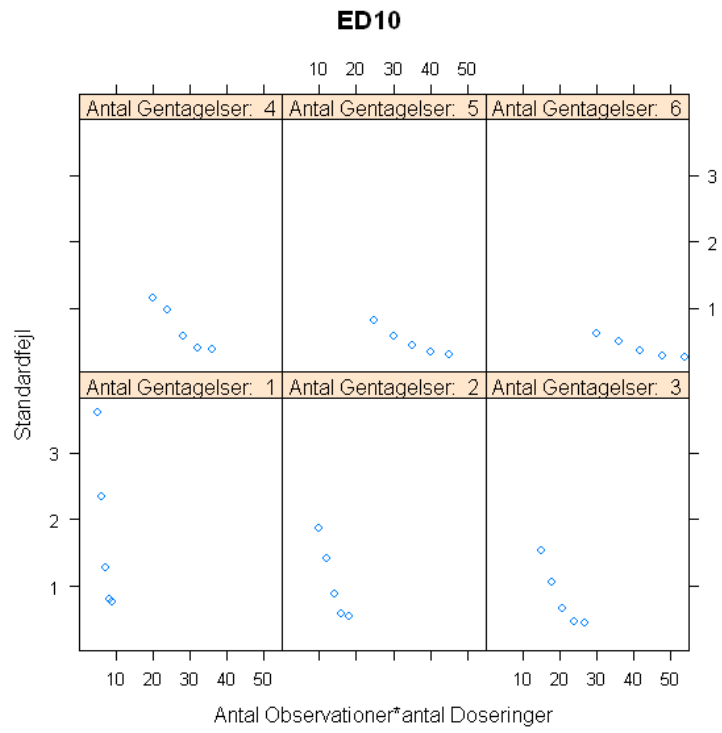
SCENARIO	EC10	EC50	EC90
A	1,46	3,06	6,39
B	0,71	3,06	13,23

For såvel scenario A som scenario B var standardafvigelsen på målefejlen 0,52. Parametrene for scenario A er baseret på ovenstående datasæt, mens scenario B er fremkommet ved en modificering af parametrene, som spreder EC-værdierne mere ud på dosis-aksen (EC10 og EC90 ligger længere fra EC50 end i scenario A). De benyttede doseringer er som nævnt ovenfor.

Betydningen af gentagelser kontra dosering/koncentrationer med kun en gentagelse er undersøgt ved tilfældig udtrækning af doseringer i den ovennævnte doseringsrække. Hvis for eksempel fem doseringer blev simuleret, så blev de fem doseringer trukket tilfældigt (uden tilbagelægning) ud fra

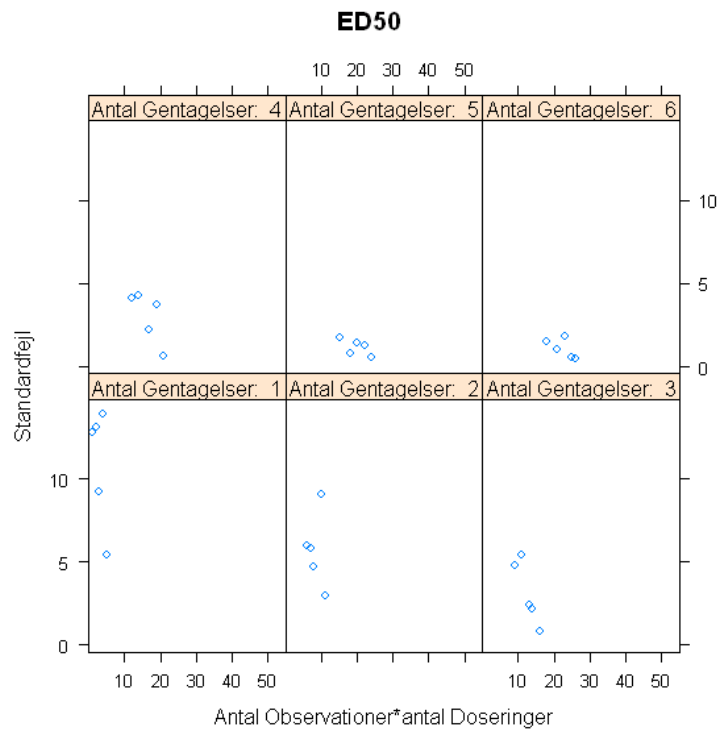
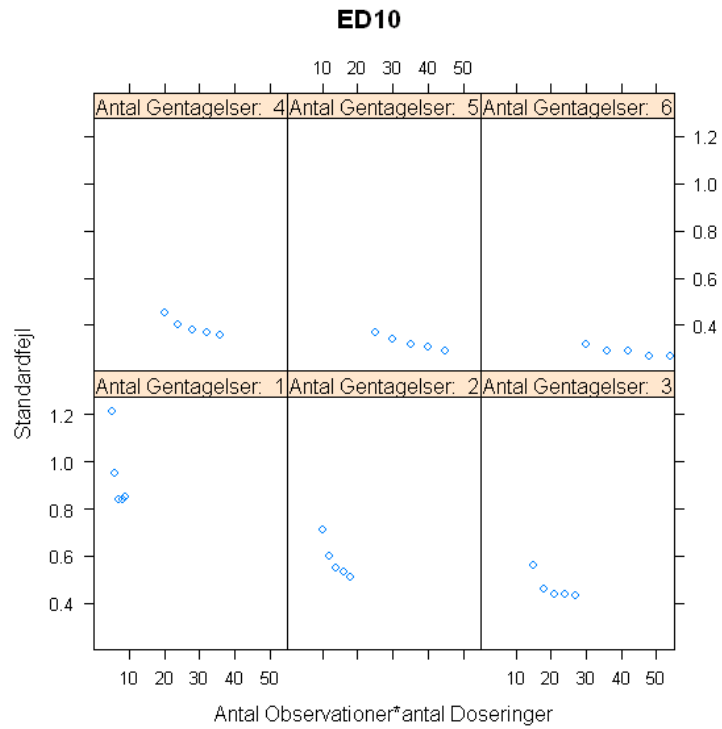
ovenstående doseringer. På tilsvarende vis blev de øvrige doseringer udtrukket (fra 6-9 doseringer). Der findes flere andre måder at udvælge doseringerne på, f.eks. kunne man starte med endepunkterne og så efterfølgende udvælge de resterende doseringer baseret på passende fraktiler. Hvis f.eks. 3 doseringer skal udvælges, så bliver den tredje valgt som medianen i fordelingen af doseringer.

For hvert datasæt er standardfejlen på EC10 og EC50 udregnet og efterfølgende blev den empiriske standardafvigelse over de 1000 datasæt fundet for hver af EC10 og EC50. I Figur 4.1 er parametrene for Scenario A benyttet og i Figur 4.2 er parametrene for Scenario B benyttet. Forskellen mellem scenario A og B er, at i B er der større afstand mellem EC10, EC50 og EC90 end i A.



Figur 4.1. Simulerede standardfej ved forskellige kombinationer af gentagelser og doseringer for Scenario A i Tabel 4.1.





Figur 4.2. Simulerede standardfejil ved forskellige kombinationer af gentagelser og doseringer for Scenario B i Tabel 4.1.

#### 4.2.4 Sammenfattende bemærkninger

For scenario A finder vi for EC10, at 9 doseringer med 1, 2 og 3 replikater per dosering (9, 18, 27 observationer) giver betydelig mindre standardfejl (ned til en halvering) end 5 doseringer med henholdsvis 2, 4 og 6 replikater (10, 20, 30 observationer). Billedet er det samme for EC50, idet standardfejlen er betydeligt mindre for 9 doseringer med 1, 2 og 3 replikater per dosering (9, 18 og 27 observationer) end for 5 doseringer med henholdsvis 2, 4 og 6 replikater (10, 20 og 30 observationer). For scenario B er billedet det samme, om end forskellen mellem standardfejlen er knap så stor.

Sammenfattende kan det konkluderes, at den højeste præcision på EC10 og EC50, når der er restriktion på det samlede antal observationer, opnås ved at bruge flere doseringer frem for at bruge flere replikater (i nogle tilfælde opnås en ganske betydelig forbedring af præcisionen).

For at kunne belyse denne problemstilling er der blevet udviklet en funktion til udvidelsespakken *drc* til **R**, således at det er muligt ud fra et sæt af modelparametre (f.eks. fra et tidligere studie svarende til det nuværende eksperiment eller baseret på værdier fra litteraturen) samt et sæt af doseringer at få beregnet en tabel, hvor simulerede, gennemsnitlige standardfejl for valgte ED-værdier bliver angivet for forskellige antal replikater.

##### 4.2.4.1 Eksempel med **R**

Vi ønsker at simulere en dosisresponskurve med en fire-parameter log-logistisk model (ligning 2.1.a) og vi antager (f.eks. ud fra tidligere forsøg) at modellen kan beskrives med parametrene:  $b=3$ ,  $c=1$ ,  $d=8$ ,  $e=EC50=10$ . Desuden antages en residual standardfejl på 0,5.

Som udgangspunkt ønsker vi at bruge følgende doseringer mellem 0, 64, 2 og 32, men med mulighed for at tilføje doseringerne 4,16 og 8 (i den nævnte rækkefølge) en for en. Disse valg specificeres på følgende måde ved at angive en række argumenter i **R**-funktionen *simDR*:

```
simDR(c(3, 1, 8, 10), 0.5, LL.4(), c(0, 64, 2, 32, 4, 16, 8),  
noSim=1000, noRep=6)
```

Beregningerne er baseret på 1000 simulationer for ED10 og ED50, fra 1 og op til 6 replikater. Resultatet er nedenstående udskrift i **R**:

```
Average 'se' based on 1000 simulations
```

```
Concentrations used: 0 64 2 32 4 16 8
```

```
EC value considered 10 (TRUE: 4.81)  
      1      2      3      4      5      6  
4  6.565232 6.182361 4.938106 3.7430477 2.3500168 2.5023711  
16 4.503998 3.635206 2.341566 3.2517200 1.6776820 1.8352134  
8  2.146454 2.161131 1.354035 0.8917755 0.7681428 0.6890755
```

```
EC value considered 50 (TRUE: 10)  
      1      2      3      4      5      6  
4  8.349600 24.492790 6.126075 4.891222 3.9353017 4.3904793  
16 2.575200  3.782776 2.525893 2.080585 1.4595111 1.3772345  
8  1.602380  1.838767 1.244030 0.636938 0.6240596 0.8164947
```

Ovenstående udskrift viser de gennemsnitlige standardfejl på EC10 og EC50 baseret på 1000 simulationer fra en log-logistisk model med parametre  $b=3$ ,  $c=1$  (nedre grænse),  $d=8$  (øvre grænse) og  $e=EC50=10$ . I hvert simulationstrin estimeres  $3 \cdot 6 = 18$  modeller: 6 modeller med 5 doseringer (0, 2, 4, 32 og 64) med antal replikater fra 1 til 6, 6 modeller med doseringer (0, 2, 4, 16, 32 og 64) med antal replikater fra 1 til 6 og endelig 6 modeller med 7 doseringer (0, 2, 4, 8, 16, 32, 64) med antal replikater fra 1 til 6. For hver estimeret model i hvert simulationstrin udregnes standardfejlen for EC10 og EC50. Gennemsnit over de 1000 simulations for EC10 og EC50 for hver kombination af antal doseringer (fra 5 til 7) og antal replikater (fra 1 til 6) er vist i udskriftet.

Igen ses det, at for det valgte scenario opnås en betydelig bedre præcision ved at tilføje flere doseringer i forhold til at bruge flere replikater. F.eks. er der for EC50 mere end en faktor 10 til forskel på at bruge alle 7 doseringer en gang i forhold til blot at bruge de 5 doseringer: 0, 2, 4, 32, 64 to gange hver.

#### 4.3 Modelvalgets indflydelse på estimater for EC-værdier

##### 4.3.1 Problemstilling

I det følgende undersøges om valget af dosis-respons model har afgørende indflydelse på estimerede EC-værdier.

##### 4.3.2 datakilde

Data stammer fra det samme eksperiment hvor den kombinerede effekt af forskellige fenoler på rajgræs (*Lolium perenne L*) blev undersøgt. Data bruges til at estimere de parameterverdier som alle øvrige datasæt simuleres ud fra. Igen betragtes et eksempel med kontinuert respons.

##### 4.3.3 Simulationsstudie

Der blev genereret 1000 datasæt baseret på hver af følgende modeller: log-logistisk, Weibull-1 og Weibull-2 (Seber & Wild, 1989, pp. 338–339). For alle 3 modeller blev følgende doseringer benyttet:

0,00; 0,94; 1,88; 3,75; 7,50; 15,00; 30,00

med 6 replikater for dosering 0 og 3 replikater for de øvrige doseringer.

For den symmetriske log-logistiske model blev følgende parameterverdier benyttet til at generere datasættene:

$$b = 2,98 \quad c = 0,48 \quad d = 7,79 \quad EC50 = 3,06$$

med standardafvigelse på målefejlen på 0,52.

For den asymmetriske Weibull-1 model (langsomt aftagende fra den øvre grænse) blev følgende parameterverdier benyttet:

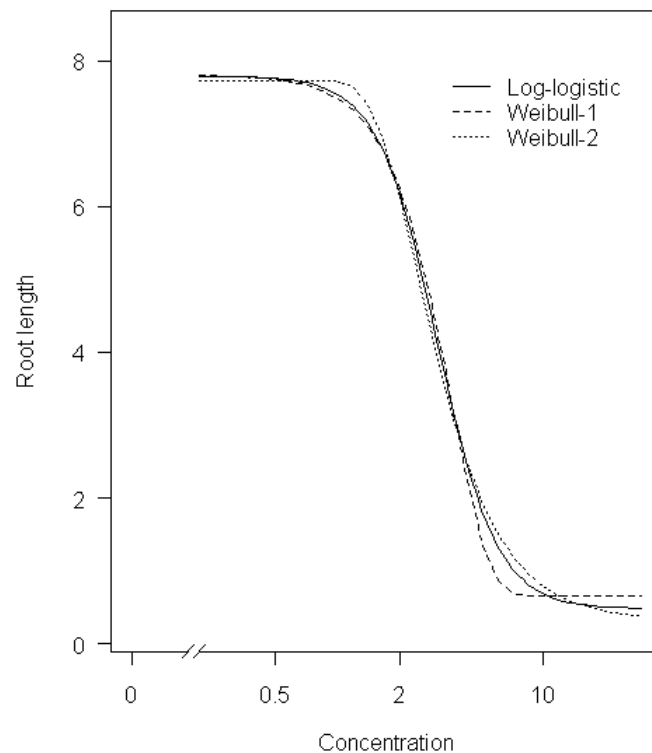
$$b = 2,39 \quad c = 0,66 \quad d = 7,81 \quad e = 3,60$$

med standardafvigelse på målefejlen på 0,55. Bemærk at parameteren  $e$  i Weibull-1 ovenfor og Weibull-2 nedenfor er vendetangenten, som netop i de to tilfælde ikke er midt på kurven

Endelig blev der for den asymmetriske Weibull-2 model (hurtigt aftagende fra den øvre grænse) benyttet parameterverdierne:

$$b = -1,97 \quad c = 0,32 \quad d = 7,73 \quad e = 2,49$$

standardafvigelse på målefejlen var på 0,51. Parametrene  $c$  og  $d$  svarer i alle tre modeller til henholdsvis den nedre og øvre grænse. Figur 4.3 illustrerer forskellene mellem modellerne (kurverne er tegnet ud fra de ovenfor angivne parametre).



Figur 4. 3. Illustration af de tre 4-parameter-dosis-responsmodeller: den Log-logistiske model (fuldt optrukne kurve), Weibull-1-modellen (lange stiplede kurve) og Weibull-2-modellen (korte stiplede kurve).

Tabel 4.1 viser absolut og relativ bias, som er den gennemsnitlige differens mellem den estimerede EC-værdi og den rigtige EC-værdi (for det relative bias er differensen endvidere blevet divideret med den rigtige EC-værdi), og mean square error (MSE), dvs. den gennemsnitlige (kvadratiske) afstand fra den estimerede EC-værdi og til den rigtige EC-værdi for de tre modeller.

Tabel 4.1. Absolut og relativ bias samt "Mean Square Error" MSE (i parentes) er angivet under hinanden i hver celle for de tre modeller: Først vises resultater, hvor der er simuleret fra den log-logistiske model, dernæst vises resultater baseret på Weibull 1-modelen og til sidst resultater fra Weibull 2-modellen.

				ESTIMERET MODEL		
				LOG-LOGISTISK	WEIBULL 1	WEIBULL 2
SIMULERET MODEL	LOG-LOGISTISK	EC10 1,46	ABSOLUT BIAS RELATIV BIAS MSE	0,088 0,060 0,21	-0,033 -0,023 0,26	0,26 0,18 0,30
		EC50 3,06	ABSOLUT BIAS RELATIV BIAS MSE	0,064 0,021 0,16	0,12 0,04 0,16	0,022 0,0071 0,23
	WEIBULL 1	EC10 1,41	ABSOLUT BIAS RELATIV BIAS MSE	0,24 0,17 0,30	0,09 0,064 0,27	0,43 0,30 0,50
		EC50 3,09	ABSOLUT BIAS RELATIV BIAS MSE	-0,0089 -0,0029 0,14	0,053 0,017 0,12	-0,079 -0,025 0,19
	WEIBULL 2	EC10 1,63	ABSOLUT BIAS RELATIV BIAS MSE	-0,084 -0,052 0,18	-0,17 -0,10 0,29	0,066 0,041 0,18
		EC50 3,00	ABSOLUT BIAS RELATIV BIAS MSE	0,13 0,045 0,19	0,29 0,067 0,20	0,10 0,033 0,27

#### 4.3.4 Sammenfattende bemærkninger

Ud fra Tabel 4.1 kan vi konstatere, at der ikke ser ud til at være store forskelle på de estimerede EC50-værdier, uanset hvilken model der blev brugt til at generere simulerede data. Bias i estimater bliver lidt større for EC10, hvis den forkerte model bruges og det samme vil gøre sig gældende for EC90 (data ikke vist).



# 5 Analyse af kontinuert respons

Inden for såvel human- og økotoxikologi udfører man ofte forsøg med et kontinuert respons, f.eks. længde, højde, vægt, absorbans etc. I modsætning til bioassays med binomiale respons som f.eks. målinger af dødelighed, hvor kun to udfald er mulige, f.eks. død/levende og mobil/immobil.

## 5.1 Variansheterogenitet

### 5.1.1 Problemstilling

Forudsætningen for at benytte en konventionel regressionsmodel, det være sig lineær eller ikke-lineær, er at vi kan beskrive kurveforløbet og at variationen i responset er konstant uanset responsniveau, og at responsen er normalfordelt. Konstant varians er i praksis kun tilnærmelsesvis opfyldt, især hvis der er store forskelle i responsniveauerne. Hvis variansen ikke er konstant udviser data variansheterogenitet.

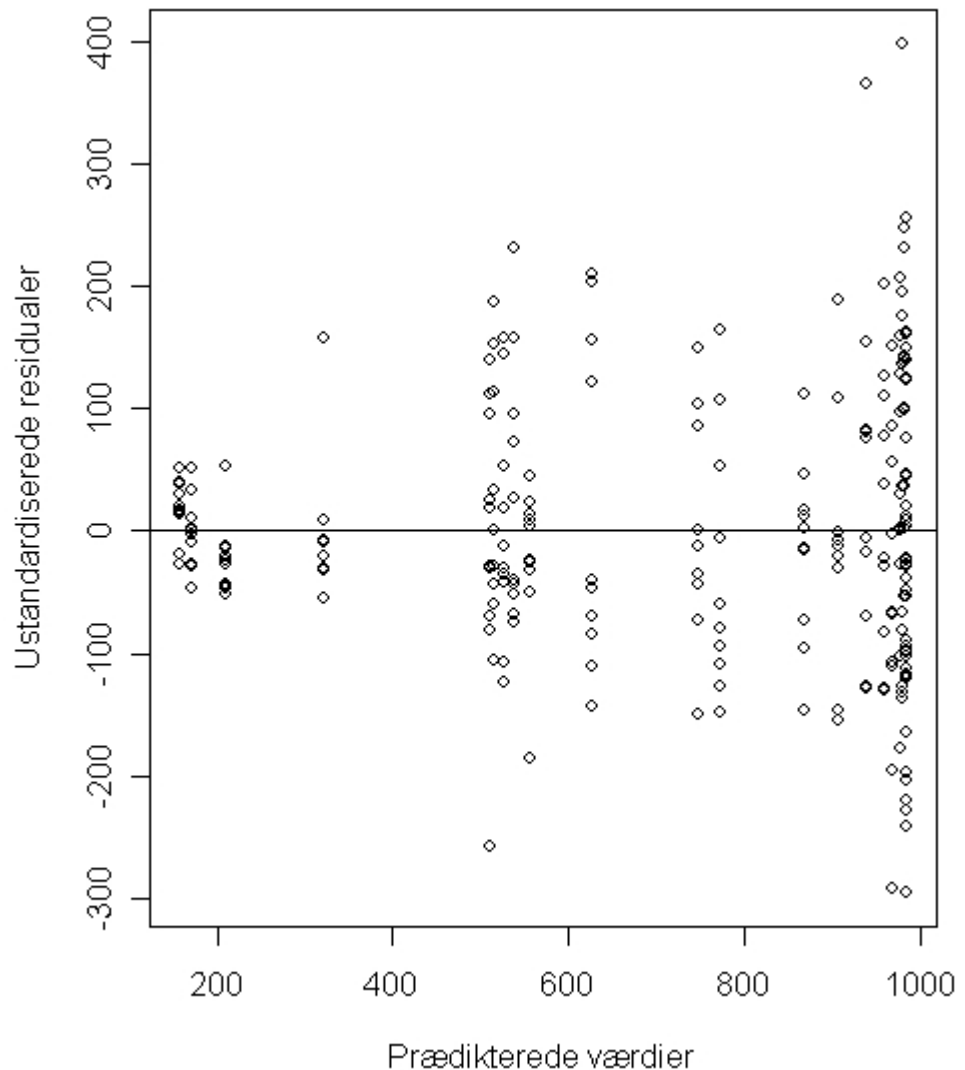
Typisk vil parameterestimerer og dermed estimerede EC-værdier ikke ændre sig meget, uanset om der tages højde for variansheterogenitet eller ej (Carroll, 2003). Nedenfor vil vi vise, hvorledes man undersøger for variansheterogenitet, og hvad man kan gøre for at afhjælpe den og vise, at ignorering af variansheterogenitet medfører forkert præcision på estimerede EC-værdier?

### 5.1.2 Datakilde

Vi vil illustrere problemet ved at anvende data fra et forsøg med små planter af *Galium aparine*, dyrket i potter i drivhus. Planterne blev sprøjtet med herbicidet phenmedipham, enten alene (herefter benævnt Herbicid 1) eller i en blanding med en ester af oliesyre (Herbicid 2). Planterne stod 14 dage i drivhuset efter herbicidbehandlingen, inden tørvægt per potte blev målt (Cabanne *et al.*, 1999).

Først betragter vi en dosis-respons-analyse af datasættet med en fire-parameter log-logistisk model, hvor variansheterogenitet ignoreres. Residualplottet fra modellen uden transformation i Figur 5.1 viser tydeligt, at der er variansheterogenitet. Det vifteformede mønster, hvor spredningen i residualerne er lille ved små prædikterede værdier og stor ved store prædikterede værdier; et typisk billede for denne type data. For at undgå variansheterogenitet kan man enten vægte eller transformere responset. Tidligere var vægtning udbredt (Finney, 1978). Det er vigtigt her at pointere, at vægtning udelukkende drejer sig om at vægte for at opnå varianshomogenitet, og at det ikke ændrer på den funktionelle sammenhæng mellem dosis og respons. En anden måde at klare variansheterogenitet på er at transformere vha. en potens eller en logaritme-funktion. Vi benytter den såkaldte Box-Cox transformation af såvel højre som venstre side af modelleringen (Streibig *et al.*, 1993). Men andre ord, ved at transformere begge sider af regressionsmodellen forbliver parameterestimererne i den oprindelige skala, og den matematiske relation mellem dosis og respons er uændret. Hvis der forekommer negative værdier af responsen, så giver det

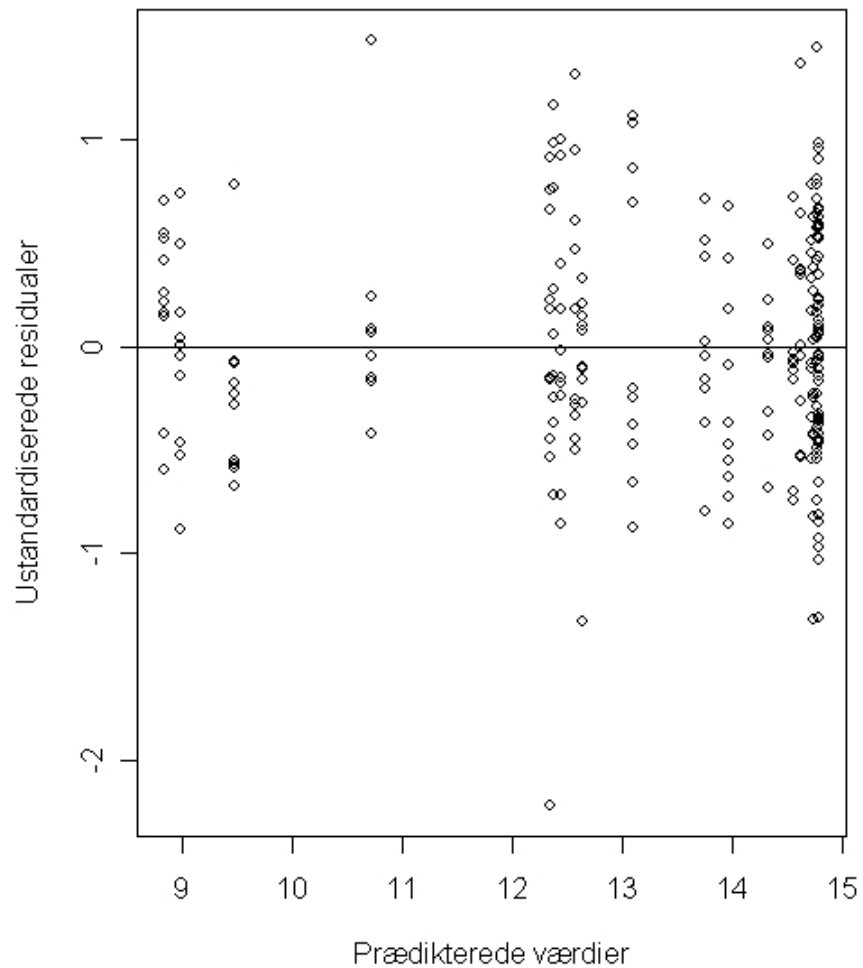
ikke mening at transformere med en potens- eller logaritmefunktion. Man kan derimod benytte Box-Cox-transformationen efter at have lagt et passende tal til alle responsværdier; det er dog ikke aktuelt for det datasæt, vi betragter. Desuden har det vist sig, at ved at benytte Box-Cox transformationen på begge sider af modellen, vil der ofte blive justeret for såvel variansheterogenitet som afvigelser fra normalfordeling. Box-Cox-transformationen er en indbygget facilitet i pakken **drc**. Den optimale Box-Cox-transformation bestemmes ud fra data.



Figur 5.1. Residualplot baseret på model uden transformation.

Benyttes den samme model, men med transformation af både respons og modelfunktion (den optimale transformation er en potensfunktion med eksponent 0,2) fås residualplottet i Figur 5.2.





Figur 5.2. Residualplot baseret på model med transformation.

Residualplottet viser ikke længere nogen systematisk vifteform som i Figur 5.1, og vi kan slutte, at modellen med transformationen har medført varianshomogenitet.

Tabel 5.1. EC10 og EC50 estimeret ud fra oprindelige data og transformerede data (Box-Cox-transformation) for to herbicidformuleringer. Standardfejl er angivet i parentes.

EC-niveau	Herbicide 1	
	Oprindelige data	Transformerede data
10	13,82 (4,32)	13,00 (3,77)
50:	52,55 (7,61)	50,82 (7,89)
	Herbicide 2	
10:	30,74 (4,29)	26,64 (4,09)
50	94,15 (5,74)	93,43 (8,11)

For Herbicide 1 ser vi, at ved at anvende transformationen får vi en standardfejl på EC50-estimatet, som er noget større (forøgelse på ca.15%) end fra modellen uden transformation. Standardfejlen på EC10-estimatet bliver en lille smule mindre med Box-Cox transformationen end for modellen uden transformation. For Herbicide 2 er billedet det samme, en forøgelse af standardfejl på EC50-estimatet og en reduktion af standardfejl på EC10-estimatet (reduktion på ca. 29%).

### 5.1.3 Sammenfattende bemærkninger

Justering for variansheterogenitet med en Box-Cox transformation kan resultere i ganske betydelige ændringer i præcisionen (mellem 15%-29%). Ikke nok med at man får en misvisende standardfejl på regressionsparametrene og dermed på EC-værdier, hvis der er variansheterogenitet, men de sædvanlige tests vil også give misvisende resultater.

## 5.2 Justering for variation fra kurve til kurve

### 5.2.1 Problemstilling

Vi betragter et dosis-responseeksperiment, som blev udført for at undersøge, om der er hormonforstyrrende effekter af et stof. Nogle gange er man pga. praktiske begrænsninger nødt til at gentage det samme eksperiment flere gange over tid, og dermed melder spørgsmålet sig, om hvorledes man kan beskrive variationen indenfor og mellem forskellige uafhængige eksperimenter af samme stof i samme biologiske system.

### 5.2.2 Datakilde

Data er luminiscens-målinger på celler som funktion af vinclozolin (doseringsområde: 0,025-3,1  $\mu\text{M}$ ). Forsøget er gentaget 6 gange. Forskelle fra forsøg til forsøg kan forventes, idet der kan være forskel på dage, laboranter, temperatur og en række andre parametre som ændrer sig fra forsøg til forsøg. I Figur 5.3 kan man se data som er dosis-responskurver med et mere eller mindre forskelligt udseende, men dog især med forskellige øvre grænser.

Vi benytter en tre-parameter log-logistisk model og antager dermed *a priori*, at den nedre grænse er 0. Først analyseres datasættet, som om der var uafhængighed mellem målinger, både inden for en kurve og på tværs af kurverne (data slås sammen henover forsøgene og én dosis-respons-kurve bliver tilpasset). Det giver følgende estimerer med standardfejl i parentes:

Hældning (b):	0,48 (0,11)
Øvre grænse (d):	1967 (161)
EC50:	0,08 (0,04)

med en estimeret residualstandardfejl på 396 (50 frihedsgrader). Bemærk at standardfejlen på EC50 er så stor at det er lige før EC50 ikke er forskellig fra 0.

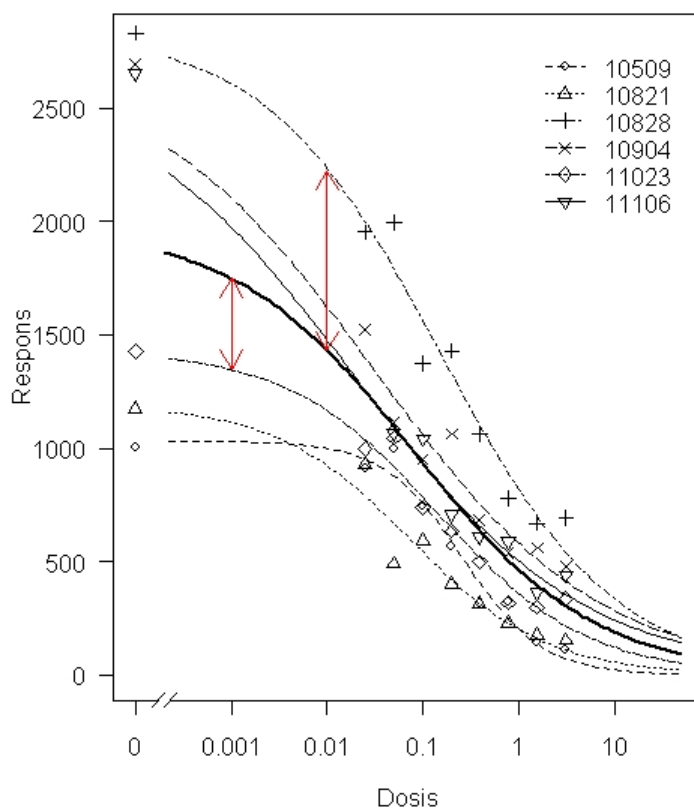
En analyse, der reducerer effekten af forsøgsspecifikke forskelle, kan baseres på en ikke-lineær regressionsmodel med tilfældige virkninger (Pinheiro & Bates, 2000). Vi kan derfor bruge en ikke-lineær regressionsmodel med tilfældige forsøgsspecifikke effekter knyttet til den øvre grænse, ifølge Figur 5.3 er den parameter (d), der formentlig har de største udsving fra et forsøg til et andet. Ved at benytte denne udvidelse af ovenstående model, fås følgende resultater:

Hældning (b):	0,47 (0,04)
Øvre grænse (d):	1976 (299)
EC50:	0.08 (0,02)

med en estimeret standardfejl for de tilfældige effekter (forsøg-til-forsøg variationen) og residualledet på henholdsvis 695 og 151. Sammenlignet med ovennævnte estimerede residualstandardfejl på 396 fra modellen uden tilfældige effekter kan man se, at modellen med tilfældige effekter har trukket mere variation ud, fordi det har været muligt at tage højde for den store variation fra forsøg til forsøg. Det er afspejlet i estimatet på 695, der er betydeligt større end residualstandardfejlen på 151. Og sidst men ikke mindst det afspejler sig i en halvering af standardfejlen for EC50.

Et statistisk likelihood ratio-test (Pinheiro & Bates, 2000, s. 83) viser, at modellen med tilfældige virkninger er signifikant bedre end modellen uden tilfældige effekter ( $p$ -værdi  $< 0,0001$ ). Det ses ligeledes, at selve estimerterne næsten ikke ændrer sig, men derimod er der store forskelle på de estimerede standardfejl. F.eks. halveres usikkerheden på den estimerede EC50 ved at bruge modellen med tilfældige virkninger.

I Figur 5.3 ses den estimerede dosis-respons-kurve baseret på modellen med tilfældige virkninger samt individuelle (forsøgsspecifikke) estimerede kurver. Eksempelvis viser de 2 lodrette, røde pile forskydninger fra det gennemsnitlige overordnede responsniveau og til 2 af de individuelle forsøg; i det ene tilfælde er der tale om en forskydning opad svarende til et større responsniveau end det gennemsnitlige, mens det i det andet tilfælde er tale om en forskydning nedad. Disse 2 forskydninger samt de 4 øvrige forskydninger (som ikke er illustreret i Figur 5.3) forklarer modellen vha. de forsøgsspecifikke tilfældige effekter.



Figur 5.3. Estimeret dosis-responskurve baseret på modellen med tilfældige virkninger (den tykke sorte linje). De individuelle forsøgsspecifikke kurver er også indtegnet. For 2 af de individuelle kurver er forskydningerne fra det gennemsnitlige niveau vist med lodrette, røde pile.

### 5.2.3 Sammenfattende bemærkninger

Vi har analyseret ovennævnte datasæt vha. to forskellige statistiske modeller: En model hvor vi så bort fra at data kom fra seks forskellige forsøg, og en model hvor der også blev taget højde for variation fra forsøg til forsøg vha. tilfældige effekter. Modellen med tilfældige effekter gav et klart bedre resultat i form af en reduceret standardfejl på EC50.

Der er yderligere den fordel, at alle kurver bliver analyseret på en gang, og der er således ikke behov for at estimere individuelle kurver for hvert forsøg, hvorfra estimater så efterfølgende skal kombineres. Modeller med tilfældige effekter giver mulighed for at opsummere flere kurver, der svarer til samme behandling, men som varierer pga. eksperimentelle forskelle. En af ulemperne er, at sådanne ikke-lineære regressionsmodeller med tilfældige effekter kan være svære at få til at konvergere (at få estimationsproceduren til at virke). Hvis det ikke lykkedes at opnå konvergens for en model med tilfældige effekter, så tyder det ofte på, at der ikke er information nok i data til at kunne anvende en sådan model. I sådanne tilfælde må man gå tilbage til modellen uden tilfældige virkninger. Modeller med tilfældige virkninger kan også benyttes til at tage højde for gentagne målinger på samme organisme.

# 6 Analyse af binomialt respons

Et binomialt respons defineres som en sum af en række uafhængige eksperimenter med kun 2 udfald, 0 eller 1, f.eks. død/levende, ikke-mobil/mobil, aktiv/ikke-aktiv.

## 6.1 Problemstilling

Skal man tage hensyn til korrelation introduceret via forsøgsdesignet, f.eks. når flere organismer af praktiske hensyn er grupperet eller samlet i en beholder? Med andre ord drejer det sig om analyse af binomialt respons med grupperede data, som er en særdeles udbredt måde at udføre bioassay på. For binomialfordelte responser forudsættes det normalt, at dosis-responsmodellen, dvs. sandsynligheden for positivt respons er 0 for dosis 0, og at den vokser asymptotisk mod 1, når dosis går mod uendelig (eller omvendt). Derfor kan man med standardmetoder få misvisende resultater, hvis man ikke kan benytte mere generelle modeller med nedre responsegrænse  $> 0$  (eller øvre grænse  $< 1$ ).

I **R**-pakken **drc** kan man vælge modeller, hvor såvel nedre som øvre grænse estimeres ud fra data.

### 6.1.1 Datakilde

Data er resultat af en larveudviklingstest med saltvandskrebsdyret *Acartia tonsa*. Disse dyr undergår morfologiske ændringer fra en række larvestadier indtil voksenstadierne nås. I larvestadierne benævnes dyrene som nauplier og i voksenstadierne som copepoditter. Der er derfor 2 mulige udfald for hvert dyr, og antal dyr, som efter en given påvirkning er i copepodit-stadiet, kan opfattes som binomialfordelt (Ole Kusk, ikke publiceret).

Da dyrene i hvert replikat er grupperet i et glas, bør man tage dette i betragtning, fordi den grundlæggende binomialfordelingsmodel bygger på antagelsen om at de enkelte dyr reagerer på stimulus uafhængigt af hinanden, en forudsætning som i dette tilfælde måske er tvivlsom, når larverne lever tæt sammen. Den type korrelation mellem individer vil ofte manifestere sig i form af overspredning, dvs. at der forekommer en større variation i data, end hvad man ville forvente ud fra den valgte model.

### 6.1.2 Analyse af data

Først betragter vi en model, som antager at der er binomialfordelt respons, som beskrives ved en 2-parameter log-logistisk dosis-respons-kurve med nedre grænse 0 og øvre grænse 1. Bemærk at det her er den log-logistiske model som vist i ligning 2.1 med den forskel at den øvre grænse er fastsat til 1. Mere præcist er modellen for sandsynligheden  $p(x)$  for at en larve bliver copepodit ved en bestemt dosis givet ved:

$$p(x) = \frac{1}{1 + \left(\frac{x}{EC50}\right)^b} \quad (6.1)$$

med parametre **b** og **EC50**. Parameteren **b** er proportional med hældning, når dosis er lig **EC50**. Denne model er identisk med en standard logistisk regressionsmodel med logaritme-transformerede doseringer (det er blot en anden parametrisering som er benyttet):

$$\text{logit}(p) = \alpha + \beta \log(x) \quad (6.2)$$

hvor **logit(p)** er defineret som **log(p/(1-p))**. Modellen baseret på Ligning (6.1) antager, at der er sandsynlighed 0 for at blive copepodit til dosis 0, og derfor er kontrolgruppen med dosis 0 overflødig og kan uden problemer udelades. Derfor er det heller ikke noget problem, at Ligning (6.2) ikke er defineret for dosis 0 (**x=0**). Ud fra modellen fås følgende estimerede EC-værdier med standardfejl i parentes:

$$\begin{aligned} EC10 &= 0,55 (0,14) \\ EC50 &= 5,83 (0,67) \end{aligned}$$

Pearson's goodness-of-fit-teststørrelsen (McCullagh & Nelder, 1989, s. 37) har en værdi på 74 og er således særdeles signifikant (p-værdi =  $1,3 \cdot 10^{-7}$ ), så modellen giver ikke nogen god beskrivelse af data. Det kan man også se på Figur 6.1 til venstre, fordi der ikke er data der tilsiger, at den øvre grænse, som vist i Ligning (6.1) er 1. Derfor er det naturligt at betragte en model, som antager, at vi har et binomialfordelt respons som beskrives ved en dosis-respons-kurve med nedre grænse 0, men en øvre grænse som skal bestemmes ud fra data. Modellen for sandsynligheden for, at en nauplie udvikles til en copepodit, er:

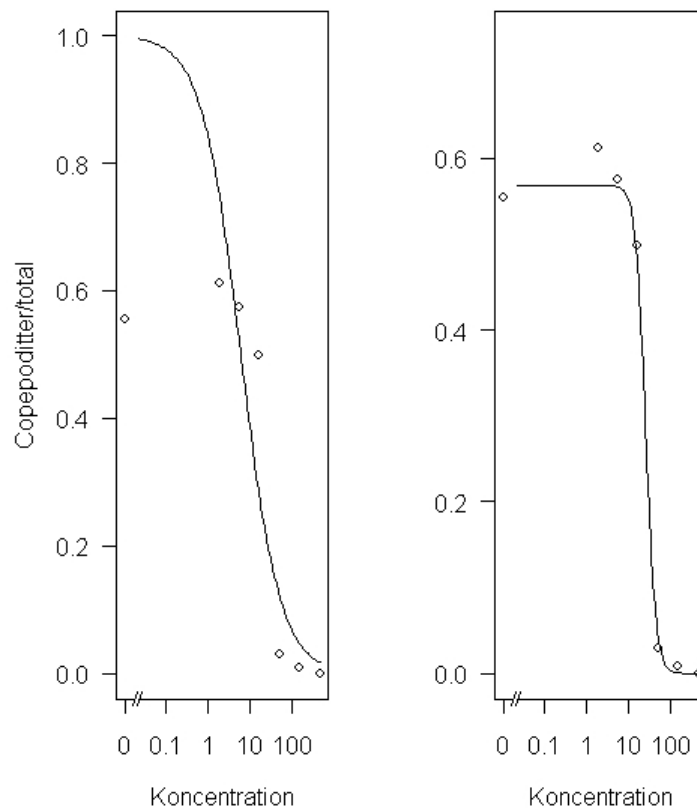
$$p(x) = \frac{d}{1 + \left(\frac{x}{EC50}\right)^b} \quad (6.3)$$

Ligning (6.3) er en tre-parameter log-logistisk model, hvor parameteren **d** betegner den øvre grænse, mens parametrene **b** og **EC50** har samme fortolkning som i ovenstående 2-parameter model i Ligning (6.1).

Denne model giver følgende estimerede EC-værdier med standardfejl i parentes:

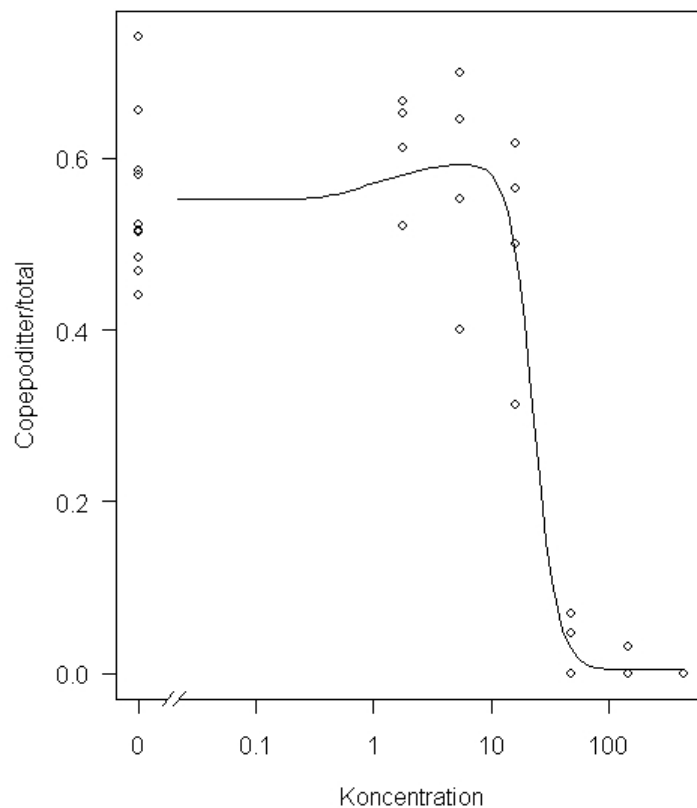
$$\begin{aligned} EC10 &= 13,60 (2,26) \\ EC50 &= 24,66 (2,57) \end{aligned}$$

Goodness-of-fit-testet er stadigvæk signifikant, dog ikke så meget som før (teststørrelse=69, p-værdi=0,0003). Den estimerede dosis-responskurve kan ses nedenfor i Figur 6.1 til højre. Vi bemærker, at de estimerede EC-værdier ændrer sig ganske meget, de bliver større end i modellen med fast øvre grænse på 1.



Figur 6.1. Estimerede dosis-responskurver baseret på model 1'er med og uden estimation af den øvre grænse for *Acartia tonsa*-datasættet.

Ud fra Figur 6.1 til højre er det tydeligt, at der er en svag stimulering ved lave koncentrationer. Dette betegnes ofte som hormesis, der kan modelleres f.eks. vha. hormesis-modellen introduceret af Cedergreen *et al.* (2005). Den resulterende estimerede dosis-responskurve er vist i Figur 6.2. Goodness-of-fit-testet er nu heller ikke så signifikant ( $p$ -værdi=0,002) som for de tidligere modeller. Rent faktisk viser denne model, at den hormesis effekt ikke er signifikant ( $p$ -værdi=0,25), og derfor kan hormesis-modellen reelt reduceres til den log-logistiske model med estimeret øvre grænse (Figur 6.1 til højre).



Figur 6.2. Det originale *Acartia tonsa*-datasæt med en estimeret dosis-responskurve baseret på hormesis-modellen introduceret i Cedergreen *et al.*, (2005). Alle datapunkter er vist.

Hermed har vi fundet en model (den log-logistiske model med estimeret øvre grænse), der beskriver middelværdistrukturen i data på en rimelig, om end ikke helt tilfredsstillende måde. Da individerne i et glas (20-45 pr glas) muligvis påvirker hinanden, er næste skridt at justere for korrelationen mellem dem. Man vil ofte forvente, at der er overspredning, fordi der er mere variation i data end hvad modellen med uafhængighedsantagelsen kan forklare. Den simpleste måde at justere på er ved at gange alle estimerede standardfejl for estimatorerne i modellen med en faktor, der afspejler hvor meget overspredning, der er til stede i data. Overspredningsfaktoren udregnes som kvadratroden af værdien af goodness-of-fit-testet, som vi allerede har brugt, divideret med antallet af frihedsgrader (McCullagh & Nelder, 1989, pp. 124-128), dvs.  $\sqrt{(69/33)}=1,45$ . Efter denne korrektion bliver resultatet for den log-logistiske model med estimeret øvre grænse:

$$\begin{aligned} EC_{10} &= 13,60 (3,27) \\ EC_{50} &= 24,66 (3,71) \end{aligned}$$

De estimerede EC-værdier er således uændrede, men de tilhørende standardfejl bliver noget større, fordi de nu bedre afspejler den reelle variation, der er til stede i de eksperimentelle data.

### 6.1.3 Afsluttende bemærkninger

Det kan være problematisk blot at bruge en standard log-logistisk regressionsmodel, som antager nedre grænse 0 og øvre grænse 1, hvor der i



virkeligheden ville være brug for at estimere den øvre grænse. Dette har den konsekvens, at EC-værdier kan blive underestimeret.

**Anbefaling:** Hvis der er tale om et forsøgsdesign, som kunne give anledning til korrelation (dvs. design hvor der er risiko for at organismerne påvirker hinanden), så bør man justere de estimerede standardfejl, så der også tages hensyn til denne ekstra variabilitet i data.

6.2 Analyse af binomialfordelt respons med modeller for kontinuert respons

### 6.2.1 Problemstilling

I litteraturen ses af og til at binomialfordelte data behandles med statistiske modeller, der forudsætter at data er kontinuerte. Spørgsmålet er derfor, hvor stor en fejl man begår ved at analysere binomialfordelte data vha. en model for kontinuerte data.

### 6.2.2 Datakilde

Vi benytter doseringer og parameterverdier estimeret baseret på *Acartia tonsa*-datasættet fra forrige afsnit.

### 6.2.3 Simulationstudie

Forsøgsvis anvender vi den tilsvarende model for et kontinuert respons, hvor øvre grænse estimeres, mens nedre grænse holdes fast på 0, se Ligning (6.3). Det gøres ved at konvertere binomialdata til procent og dermed betragte dem som kontinuerte data. Dette giver følgende estimerede EC-værdier:

$$EC_{10} = 15,14 (2,63)$$

$$EC_{50} = 24,92 (4,88)$$

Selve estimererne er ikke voldsomt forskellige fra modellen med binomialfordelt respons (med nedre grænse 0 og øvre grænse 1), men de tilhørende standardfejl ændres betragtelig. I det følgende vil vi undersøge nærmere, hvilken konsekvens det har at betragte binomial respons som om det var kontinuert.

Der blev genereret 1000 datasæt baseret på binomialfordelingen ud fra en to-parameter log-logistisk regressionsmodel med en hældning  $b$  på 0,93 og en  $EC_{50}$ -værdi på 5,83 og med samme doseringer som i *Acartia tonsa*-datasættet for hver af de følgende totale antal larver: 10, 20, 30, 40, 50, 60, 100, 500, 1000 og 10000.

For hvert datasæt blev både den korrekte binomiale model (B) og den tilsvarende kontinuerte model (K) tilpasset til data. Resultaterne i form af relativt bias, dvs forholdet mellem forskellen (differensen) fra estimat til den rigtige EC-værdi for den kontinuerte og binomialfordelingsmodellen, og efficiens, dvs. forholdet mellem standardfejlen på estimererne baseret på den kontinuerte og binomialfordelingsmodellen, er opsummeret i Tabel 6.1.

Tabel 6.1 Sammenligning af relativ Bias og efficiens for EC10 og EC50 ved at antage kontinuert normalfordeling og antage den korrekte binomialfordeling.

ANTAL DYR PER DOSIS	BIAS		EFFICIENS	
	EC10	EC50	EC10	EC50
3	1,06	1,00	1,34	1,03
5	1,02	1,00	1,23	1,03
10	1,00	1,00	1,21	1,02
20	1,00	1,00	1,22	1,02
30	1,00	1,00	1,21	1,02
40	1,00	1,00	1,21	1,02
50	1,00	1,00	1,23	1,02
60	1,00	1,00	1,22	1,02
100	1,00	1,00	1,22	1,01
500	1,00	1,00	1,19	1,01
1000	1,00	1,00	1,17	1,02
10000	1,00	1,00	1,21	1,01

Tabel 6.1 viser, at der ikke er noget nævneværdigt bias i estimerne fra den kontinuerede model i forhold til binomialfordelingsmodellen. For efficiens giver det samme billede som sammenligningen af analyserne for *Acartia tonsa*-datasættet gjorde ovenfor, nemlig at parametrene ikke ændredes sig meget (der er ingen bias), mens standardfejlen gør det. Den kontinuerede model producerer for store standardfejl på EC-værdierne. Mest markant er det for EC10 med en efficiens på omkring 1,2, mens der ikke er stor forskel på standardfejlene for EC50. De relative forskelle ændrer sig ikke meget, uanset hvor stort antal larver der bruges. Efficiensen for EC5 (ikke vist i Tabel 4) var på omkring 1,23, og altså lidt højere end for EC10.

#### 6.2.4 Afsluttende bemærkninger

Resultater viser, at uanset hvor stort antal larver der bruges, så vil den kontinuerede model give for store standardfejl på EC10-værdierne. Derimod forringes præcisionen ikke meget, når man kun betragter EC50. Disse resultater viser, at det ikke er korrekt at antage, at en kontinuert model er acceptabel, hvis blot antal larver (generelt: antal organismer) er stort nok. Dermed er vores resultater i modstrid med gængse anbefalinger, som f.eks. fremsat af Environment Canada (2005, s. 84).

## 7 "Lack of fit"-test

Når der er replikater inden for doseringer i et bioassay har vi mulighed for at sammenligne regressionsanalysen med en mere generel model (ofte en variansanalyse), dvs. udføre et lack-of-fit-test af dosis-responsmodellen. En variansanalysemodel tager ikke højde for, hvorledes den funktionelle sammenhæng imellem respons og dosis ser ud. Derimod forudsætter regressionsanalysen, som det er vist gentagne gange i denne rapport, at vi kender den funktionelle sammenhæng og at vi kan beskrive denne med en matematisk model. Hvis en sammenligning mellem regressions- og variansanalysemodellen ikke er signifikant vurderet ud fra et F-test, så betyder det, at der ikke er grund til at tro at regressionsanalysen ikke er lige så god til at beskrive variationen i data som variansanalysen (Draper and Smith 1981). Dosis-responsmodellen foretrækkes, idet den beskriver data vha. færre parametre. F-testet giver således en god indikation af kvaliteten af en regressionstilpasning. I næste afsnit betragter vi et lack-of-fit-test, der kan bruges når man ikke har replikater og derfor ikke kan anvende en variansanalyse.

### 7.1 Test baseret på kumulerede residualer

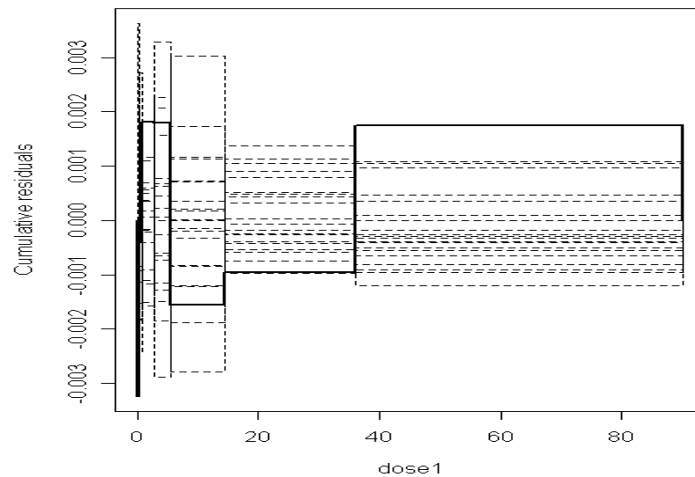
Et attraktivt alternativ til lack-of-fit-testet er testet baseret på de kumulerede residualer. Hvis de rå residualer kaldes  $e_i$  og de anvendte doseringer  $x_i$ , så er testet baseret på følgende sum:

$$W_n(x) = \sum_{i=1}^n I(x_i \leq x) e_i / \sqrt{n} \quad (7.1)$$

hvor  $I(x, x)$  er en indikatorfunktion (den er 0 hvis  $x_i$  er større end  $x$  og 1 ellers).  $W_n(x)$  kan beregnes for hver værdi  $x$  i doseringsintervallet (vi siger, at vi har en **proces** i  $x$ ). Det er således muligt at tegne processen  $W_n(x)$  op som en funktion af dosering. Et test kan baseres på maksimum for  $W_n(x)$ . Hvor stor er maksimum i forhold til hvad man ville forvente såfremt modellen er rigtig. Det er nemt at simulere fordelingen af maksimum af  $W_n(x)$ . Tests baseret på kumulerede residualer findes for lineær regression (Su & Wei, 1991), generaliseret lineær regression (Lin *et al*, 2002) og for modeller i overlevelsesanalyse (Lin *et al*, 1993). En stor fordel ved dette test er, at der ikke kræves gentagelser ved de enkelte doseringer. Testet er implementeret i **R**-pakken **drc** og i Figur 7.1 ses de kumulerede residualer som funktion af dosis.

For det betragtede datasæt (se afsnit 9, datakilde 1 for en detaljeret beskrivelse af datasættet) fås nedenstående plot af den observerede  $W_n(x)$  (den fuldt optrukne kurve). Ud over selve den observerede proces er der også indtegnet 20 simulerede processer (de stiplede linjer) for at vise hvor meget processen må variere såfremt modellen er rigtig. P-værdien baseret på 1000 simulerede processer er 0,07, som modsat p-værdien fra det ANOVA-baserede lack-of-fit-test på 0,0009, ikke tyder på nogen markant afvigelse fra modellen.

Lack-of-fit-testet bygger på den antagelse, at der er varianshomogenitet dvs. at variationen i responsen er den samme for alle doseringer. Denne antagelse kan ikke undersøges i situationer, hvor der kun er gentagelser i kontrolgruppen (dosering 0); ved vi overhovedet intet om variationen for alle de andre doseringer. I ovennævnte eksempel er variationen temmelig lille i kontrolgruppen, og på den baggrund behøver afvigelser mellem observationer og den estimerede log-logistiske model til de andre doseringer ikke at være særligt store for at kunne give et signifikant lack-of-fit-test. Hvis omvendt variationen i kontrolgruppen er rigtig stor, så vil lack-of-fit-testet resultere i en p-værdi tæt på 1, et ikke-signifikant resultat.



Figur 7.1. Kumulerede residualer er indtegnet som funktion af dosis (sort linje). De stiplede linjer svarer til 20 simulerede processer. Der ser generelt ud til at være en god overensstemmelse mellem de observerede og simulerede kumulerede residualer.

## 7.2 Afsluttende bemærkninger

Det er muligt at udføre et lack-of-fit-test, uanset om der er replikater af de anvendte doseringer. Generelt vil testene dog have lav styrke for små datasæt, men ved at øge antallet af doseringer vil nogle af testene få en ganske betydelig styrke. Dette er en af konklusionerne i Ritz & Martinussen (2008), hvor lack-of-fit-test for dosis-responsmodeller sammenlignes. En anden hovedkonklusion er, at testene ikke bliver meget bedre af, at der er flere replikater for et fast antal doseringer, men det gør de til gengæld, hvis antallet af doseringer øges.

## 8 Semi- og ikke-parametrisk dosis-respons-modellering

Et alternativ til de parametriske modeller for binomialfordelte og kontinuerte data, som vi har set på i de tidligere afsnit, er ikke-parametriske eller semi-parametriske dosis-responsmodeller, hvor dosis-responskurven beskrives vha. en blanding af en parametriske dosis-responsfunktion og en fuldstændig uspecificeret ikke-parametrisk funktion.

### 8.1 Alternative dosis-responsmodeller

Den således fremkomne model kaldes model-robust, fordi den er robust over for afvigelser fra den parametriske model (Nottingham & Birch, 2000). Modellen kan skrives som:

$$E(y_i) = \lambda f_{IP}(x_i) + (1 - \lambda) f_p(x_i, \beta) \quad (8.1)$$

hvor  $x_i$  og  $y_i$  henholdsvis betegner dosering og den tilhørende respons. Den parametriske og den ikke-parametriske komponent betegnes henholdsvis  $f_{IP}$  og  $f_p$ , mens  $\lambda$  betegner blandingsparametren der bestemmer i hvor høj grad den ikke-parametriske komponent skal bruges til at beskrive responsen. Man kan endda helt udelade den parametriske del ( $\lambda=1$ ) og dermed så få en rent ikke-parametrisk dosis-responsmodel. Den ikke-parametriske komponent kan estimeres vha. lokal lineær regression, hvor der for hver dosering lokalt tilpasses en vægtet lineær eller kvadratisk regression model (Nottingham & Birch, 2000). Da regressionen er lokal benyttes et begrænset interval af doseringer (et **vindue**) af en vis **båndbredde** i estimationen; for et givet datasæt fastsættes båndbredden ofte automatisk vha. krydsvalidering (for en lang række båndbredder undersøges det, hvordan det resulterende model fit prædikterer i en dosering (i datasættet), som ikke indgik i estimationen af modellens parametre).

Alternativt kan estimation for binomialfordelte data baseres på den med lokal regression beslægtede lokal (quasi) likelihood-metode (Loader, 1999; Park & Park, 2006). Vægtene aftager med voksende afstand fra den betragtede dosering til de øvrige doseringer i datasættet, således at doseringer tæt på den betragtede dosering får mest indflydelse på tilpasningen af regressionsmodellen. De præcise vægte udregnes vha. en vægtfunktion eller **kjerne**, som f.eks kan være eksponentielt aftagende fra 1 (for afstand 0) til 0 for en stor afstand (Cleveland, 1979; Loader, 1999; Nottingham & Birch, 2000). For begge typer modeller kan EC-værdier bestemmes vha. invers regression, dvs. ved at tilbagetransformere den responsværdi der svarer til det ønskede EC-niveau (Nottingham & Birch, 1996). En tredje ikke-parametrisk metode (for binomialfordelte data) er direkte at estimere  $EC_x$  som en funktion af  $x$  ved at kombinere lokal lineær regression med en efterfølgende estimation af en sandsynlighedstæthed (Dette **et al.**, 2004).

Betragt følgende datasæt med 6 doseringer fra en akut test (efter 24 timer) på *Acartia tonsa* (fra Ole Kusk):

<i>dosis</i>	<i>antal</i>	<i>total</i>
0,000	0	19
1,000	4	20
1,500	13	20
2,250	18	21
3,375	20	21
5,060	22	22

Estimater for EC10 og EC50 baseret på den tredje ikke-parametriske metode (Dette *et al*, 2004) med tilhørende konfidensintervaller er vist i nedenstående tekstabeltabel (en båndbredde på 0,5 er benyttet). Der er tale om bootstrap-intervaller, dvs. intervaller baseret på fordelingen af estimaterne fra estimerede modeller, hvor de permuterede residualer fra den oprindeligt estimerede model er brugt som restled. Til sammenligning er også de tilsvarende estimater baseret på en to-parameter logistisk regressionsmodel (med logaritmeret dosis) vist, hvor konfidensintervallerne er udregnet på sædvanlig vis.

	IKKE-PARAMETRISK METODE	LOGISTISK REGRESSION
EC10	0,50 [0,01; 0,99]	0,79 [0,65; 0,93]
EC50	1,44 [0,82; 2,05]	1,34 [1,25; 1,44]

Af tabellen ses det, at estimater ikke er voldsomt forskellige fra dem der er bestemt ved den parametriske model, men konfidensintervaller for den ikke-parametriske model er til gengæld noget bredere end for den parametriske model.

## 8.2 Afsluttende bemærkninger

Selvom ikke-parametriske metoder er attraktive i den forstand, at de ikke bygger på en eksplicit matematisk model for dosis-reponsrelationen, så er erfaringerne med disse metoder begrænset. Spørgsmål vedr. disse modellers effektivitet, monotonicitet og asymptotiske egenskaber relativt til gængse parametriske modeller er uafklaret.

# 9 Eksempler på dosis-respons problemstillinger

Dette afsnit giver nogle eksempler på analyser af dosis-respons eksperimenter, der indeholder en række af de problemstillinger, som er nævnt i kapitel 4-8. Som tidligere nævnt kan valget af regressionsmodel bygge på kendskab til, hvorledes virkningsmekanismen er for pågældende stof, men ofte er det erfaringerne fra tidligere eksperimenter, der bestemmer regressionsmodellens form og beliggenhed. Problemstillingerne i forhold til dataanalyse er ofte affødt af, hvilket testsystem og hvilken organisme man benytter, og sidst men ikke mindst hvilken responsvariabel man vælger at måle. Inden for en og samme organisme kan valg af responsvariabel have særdeles stor betydning for, hvilken dosis-responsmodel der bør benyttes.

## 9.1 Øvre og nedre asymptoter

### 9.1.1 Problemstilling

For de fleste responsvariable ligger grænseværdierne for øvre og nedre grænse i et givet interval, som er afgrænset nedad af dosis 0 og opad af en passende høj dosering. Mens respons ved dosis 0 antager en "naturlig" grænse, der kan være maksimum eller minimum afhængig af valget af respons, er det straks vanskeligere at vurdere asymptotiske grænser, når dosis går mod uendelig. Årsagerne kan være mange, f.eks. at det valgte doseringsinterval ikke er tilstrækkeligt til at beskrive hele kurveforløbet fra ingen virkning ved små doseringer til stor virkning ved store doseringer. Der kan også være problemer med opløselighed af teststoffet. Forskellen på øvre og nedre grænse kan ligeledes være meget påvirket af, hvor lang tid der går fra organismen bliver eksponeret for stoffet, og indtil man måler virkningen. Det er faktisk en problemstilling, som kun kan løses ved at kende hastigheden for udvikling af symptomer og påvirkning af det målte respons.

Manglende værdier især ved store doseringer for en monotont aftagende funktion kan medføre, at der ikke er nogen veldefineret nedre grænse, og at man derfor kan få urealistiske, negative nedre grænser for den estimerede dosis-responskurven.

### 9.1.2 Datakilde 1 (bioassay med blandet bakteriekulturer)

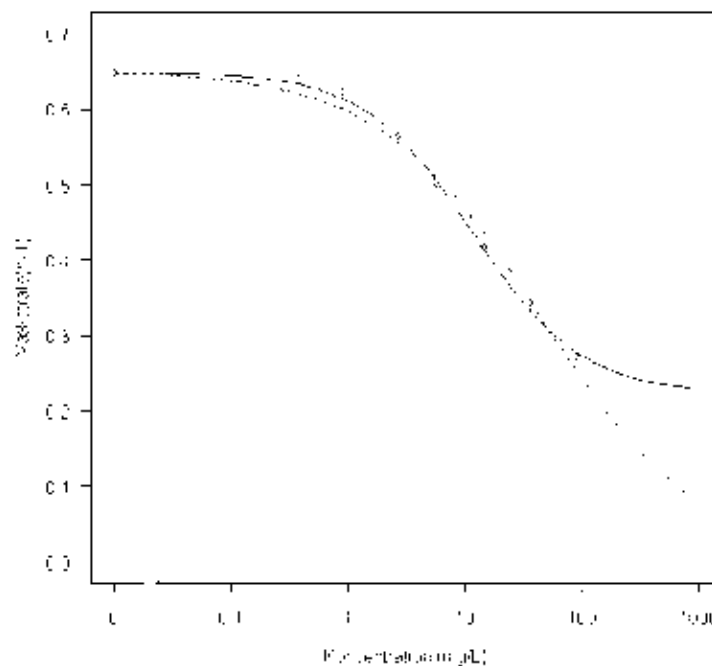
Nedenstående data stammer fra et større blandingsforsøg, hvor effekter af antibiotika undersøges overfor en blandet bakteriekultur af aktiv-slambakterier (Christensen *et al*, 2006). Forsøget er udført således, at der for hvert dosis-responsforløb er syv koncentrationer uden gentagelser og otte gentagelser for den ubehandlede kontrol. Vi betragter dosis-responsforløbet, hvor dosis er erytromycin (mg/L), og endpoint er den relative vækstrate ( $h^{-1}$ ) af biomasse i forhold til den ubehandlede kontrolgruppe, målt 4 timer efter eksponering vha. turbiditet med et spektrofotometer ved 530nm. Forsøget er udført efter ISO 15222 (ISO 15522 1999).

Som det fremgår af Figur 9.1, er der ikke data til at beskrive den nedre grænse ved høje doseringer. Ud fra biologiske betragtninger vil man forvente, at vækstraten vil falde til omkring 0 ved meget høje doseringer. Imidlertid viser en tilpasning af den log-logistiske model med såvel fri øvre som nedre grænse en anelse mere præcis beskrivelse af data end en model med en fastsat nedre grænse på 0 (vurderet ud fra Figur 9.1). Parametrene er vist i Tabel 9.1.

Tabel 9.1. Parameterestimer for den log-logistiske model med fire og tre parametre (Ligning (2.1a) og (2.1b)).

Parameter	Estimat	Standardfejl
Fire-parameter-model		
b	0,937	0,068
c	0,223	0,020
d	0,650	0,003
ED50	11,600	1,610
Tre-parameter-model		
b	0,631	0,035
d	0,652	0,005
ED50	43,896	3,762

Som det fremgår af Tabel 9.1 og Figur 9.1 er der (ikke overraskende!) store forskelle på ED50, alt efter om man benytter en model med såvel fri øvre som nedre grænse eller kun med fri øvre grænse (fire eller tre parametre i den log-logistiske model). I Tabel 9.1 kan man tydeligt se, at ED50 falder med omkring en faktor 3 ved at gå fra en fire-parameter-model til en tre-parameter-model.



Figur 9.1. Sammenligning mellem en fire-parameter log-logistisk model (fuldt optrukket linje) og en tre-parameter log-logistisk model (stiplet linje).



Som nævnt tidligere er der ingen biologiske belæg for at der skulle være en nedre grænse på kurven som er forskellig fra nul. Der biologisk set gode grunde til at tilpasse med en tre-parameter log-logistisk model, da der i forsøg med andre stoffer er fundet fuld hæmning. I nærværende forsøg har det ikke været muligt at teste højere koncentration af stoffet pga. stoffets opløselighed. Lack-of-fit-testet beskrevet i afsnit 7 vil dog i det her tilfælde bekræfte, at tre-parameter-modellen er acceptabel, så denne model vil være den simpleste biologisk og statistisk acceptable model.

### 9.1.3 Datakilde 2 (pesticider og brystcancer celler)

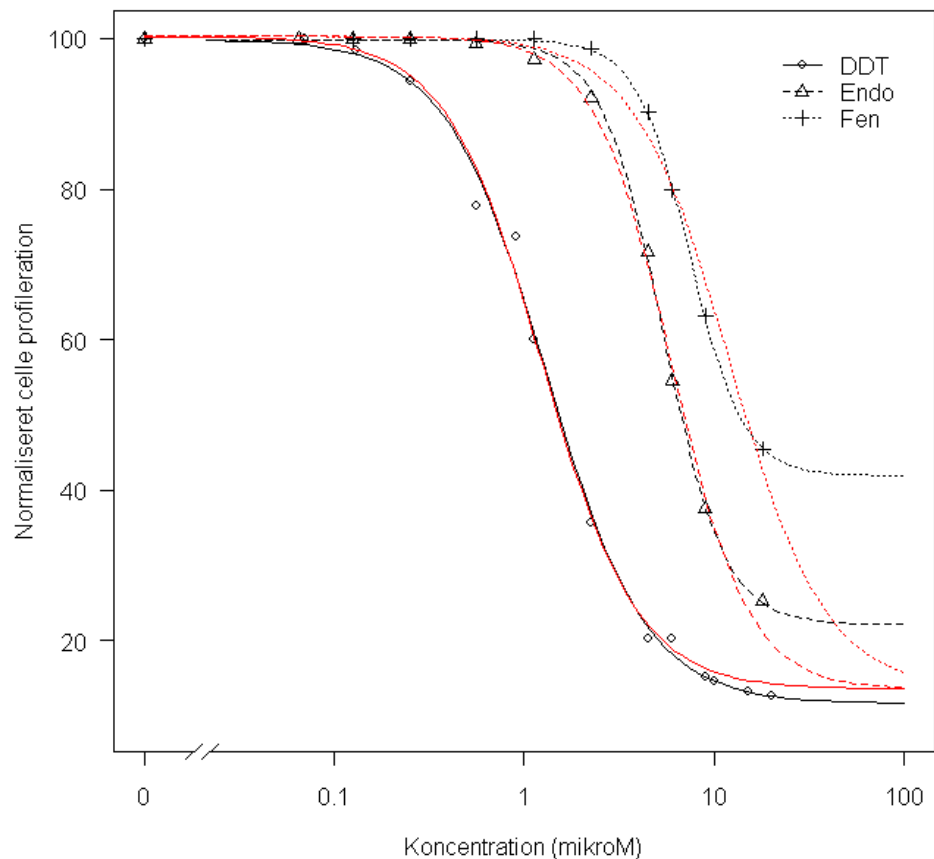
Det næste eksempel er fra et forsøg med flere dosis-responskurver og fokuserer på, hvorledes man kan slå parametre for nedre og øvre grænse sammen. Pesticiderne DDT, endosulfan (Endo) og fenarimol (Fen) er undersøgt for østrogene effekter *in vitro* i humane brystcancer celler (Vinggaard, 2003). Forsøgets formål var at undersøge toksiciteten af de tre pesticider, og derfor er den relative styrke mellem dem et godt mål for deres rangorden rent toksikologisk. Proliferation af cellerne er helt afhængig af tilstedeværelsen af østrogener. Cellerne 'udsultes' for østrogener, hvorefter et teststof tilsættes og det undersøges om celleproliferationen øges med koncentrationen af teststoffet. Endpoint er en absorbansmåling (IRPE), der indirekte er et udtryk for antal celler i brøndene. Biomassen er normaliseret vha. kontrolgruppen.

Tabel 9.2 og Figur 9.2 viser simultan tilpasning af en fire-parameter log-logistisk model til data. Det ses, at stofferne har forskellig maksimaleffekt (forskellig nedre grænse). Som det fremgår visuelt af Figur 9.2 har de tre stoffer fælles øvre grænse, fordi de allerede er skaleret mod kontrollen. Lack-of-fit-testet for modellen er ikke-signifikant.

Tabel 9.2. Parameterestimer baseret på en simultan tilpasning af en fire-parameter log-logistisk model til data (Ligning 2.1a). De tre kurver antages at have samme øvre grænse, men forskellige nedre grænser.

Parameter	Estimat	Standardfejl
b DDT	1,645	0,099
b Endo	2,685	0,251
b Fen	3,084	0,384
c DDT	11,631	1,489
c Endo	22,106	2,369
c Fen	41,915	2,901
d	99,853	0,462
ED50 DDT	1,298	0,052
ED50 Endo	5,387	0,186
ED50 Fen	7,486	0,393

Umiddelbart ser det ikke ud til i Figur 9.2, at den nedre grænse for Endo og Fen er veldefineret ud fra data. Modellen, hvor det antages, at der er fælles øvre og nedre grænse for de tre pesticider, giver imidlertid en signifikant dårligere beskrivelse af data. Og i Figur 9.2 (røde linjer) kan man se hvorfor: Der er systematiske afvigelser mellem de estimerede kurver under den model og så data for DDT og især Endo.



Figur 9.2. Estimeret fire-parameter log-logistisk dosis-responskurve for brystcancer cellers proliferation for tre forskellige pesticider. De sorte linjer svarer til en model med fælles øvre grænse og pesticid-specifik nedre grænse. De røde linjer er for en model med fælles øvre og nedre grænse for de tre pesticider.

#### 9.1.4 Afsluttende bemærkninger

Som det fremgår af de eksempler, er spørgsmålet om øvre og nedre grænser for kurver særdeles relevant, fordi de alt andet lige har stor indflydelse på de estimerede ED-værdier. Der er i tidens løb lavet mange beregninger, for at undgå netop problemet med forskellige øvre og nedre grænser (et eksempel er at relativere responsen til et kontrolniveau). Derfor er det vigtigt at overveje, om der er biologiske argumenter, der taler for at fastsætte den nedre eller øvre grænse. Desuden kunne man overveje om ED-værdier skal sammenlignes på et absolut eller relativt (eventuelt indskrænket) responsniveau (Ritz *et al.*, 2006).

### 9.2 Forskellige endpoints

#### 9.2.1 Problemstilling

Ofte vil man *a priori* have defineret et eller flere endpoints, som skal danne baggrund for en vurdering af toksiciteten. Et givet stof kan have virkninger på mere end en stofskifteproces i en plante eller et dyr. Selv om man kender virkningsstedet for et stof, kan afledte effekter være af stor betydning for at vurdere stoffers toksicitet på en organisme. Ofte vil man derfor i dosis-

responsforsøg være interesseret i at finde det endpoint, som er mest følsomt, og derfor vil man gerne måle mere end et endpoint.

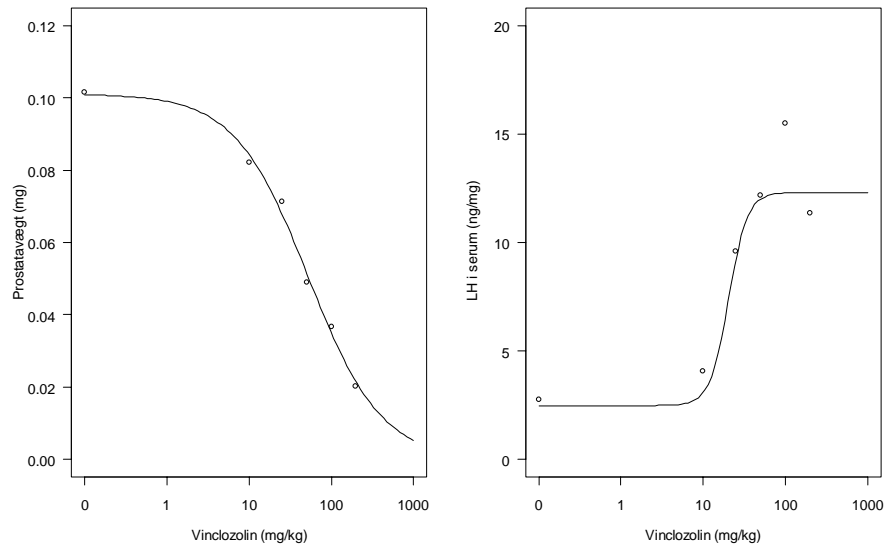
### 9.2.2 Datakilde 3

Formålet med dette dosis-responsforsøg var at bestemme det endpoint, der er mest følsomt over for et fungicid, som mistænkes for at have hormonforstyrrende virkning. I dosis-responsforsøg med fungicidet vinclozolin i kastrerede testosteron-behandlede rotter, blev der målt 2 endpoints: 1) prostatavægt og 2) koncentration af luteiniserende hormon (LH) i serum. Begge endpoints giver udslag ved 'antiandrogene effekter' og desuden er de to endpoints målt på de samme dyr. Testosteron er nødvendigt for vækst af prostata, og et antiandrogen stof som vinclozolin blokerer for testosterons effekt på prostata, idet stoffet virker som antagonist for androgenreceptoren i prostata. Sekretionen af LH fra hypofysen regulerer testosteronproduktionen i testiklerne og under fysiologiske forhold er der en negativ feedback, således at LH sekretionen øges ved for lav koncentration af testosteron og omvendt ved for højt testosteronniveau. Et antiandrogen som vinclozolin blokerer androgen receptoren i hypofysen og blokerer dermed for den negative feedback. Derved forøges LH niveauet i serum (Hass *et al.*, 2006)

I Figur 9.3 og Tabel 9.3 er der henholdsvis en illustration af den estimerede log-logistiske dosis-responsmodel sammen med gennemsnit af de oprindelige data (per dosering) og et sammendrag af parameterestimerterne for de to estimerede dosis-responskurver baseret på 2 separate analyser, en for hvert endpoint. Det var kun nødvendigt at benytte en Box-Cox-transformation til at opnå varianshomogenitet i forbindelse med LH serum responset. Lack-of-fit-testet var ikke-signifikant for begge endpoints, så den log-logistiske model giver en tilfredsstillende beskrivelse af data for de 2 endpoints.

Tabel 9.3. Sammendrag af anvendte log-logistiske regressionsmodeller for de to endpoints, prostatavægt (Ligning 2.1b) og LH serum (Ligning 2.1a).

Parameter	Estimat	Standardfejl
Prostatavægt		
b	0,979	0,108
d	0,101	0,004
ED50	52,155	6,389
LH-serum		
b	-3,843	3,714
c	2,481	1,351
d	12,306	1,673
ED50	20,417	6,944



Figur 9.3. Estimerede log-logistiske dosis-responsmodeller for hver af de 2 endpoints, prostatavægt og koncentration af LH-serum.

Som det fremgår af Tabel 9.3 er ED50 for LH-serum på 20,4 mg pr kg således knap det halve af ED-værdien for prostatavægten. Det mest følsomme endpoint er således LH-serum på ED50-niveau.

### 9.2.3 Datakilde 4

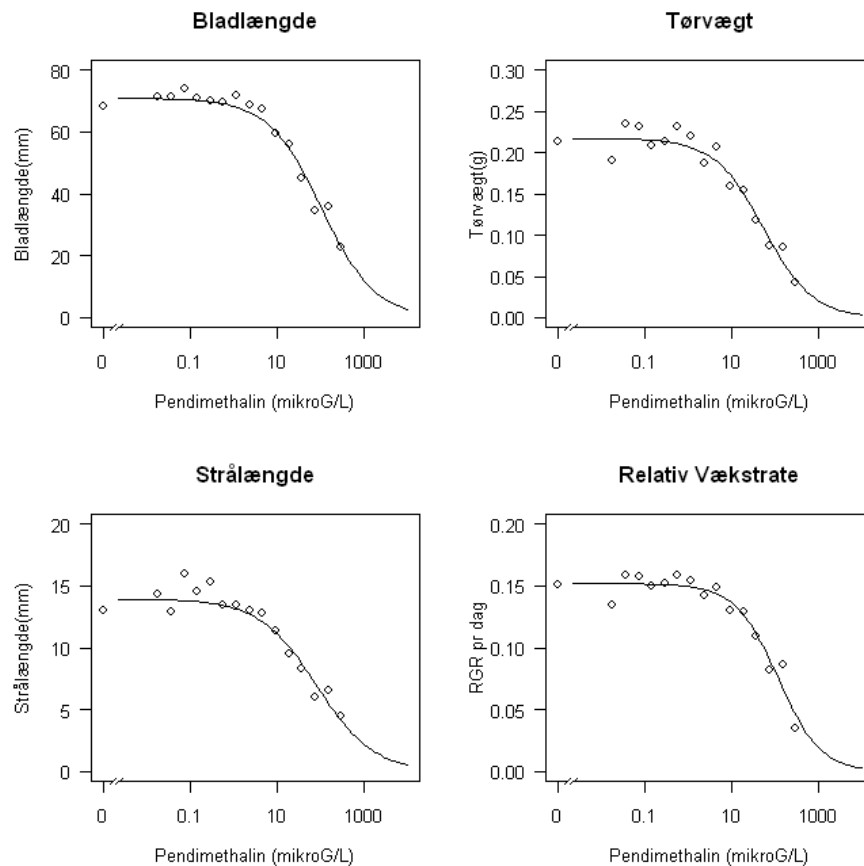
I et andet forøg med bygplanter i vandkultur, målte man forskellige endpoints med stigende koncentrationer af pendimethalin i klimakammerforsøg (Cedergreen, 2008). I dosis-responsforsøget indgik i alt 16 koncentrationer i  $\mu\text{g/L}$ , som er usædvanlig mange koncentrationer for den type forsøg.

Tabel 9.4 og Figur 9.4 viser estimerede dosis-responskurver baseret på en tre-parameter log-logistisk model anvendt på hvert endpoint separat. En Box-Cox-transformation for at stabilisere variansen var kun i et tilfælde nødvendig.

Tabel 9.4. Sammendrag af dosis-responsanalyser (ligning 2.1b) for forskellige endpoints i forsøg med byg i vandkultur med stigende koncentrationer af pendimethalin ( $\mu\text{g/L}$ ). For strållængde var det nødvendigt at bruge en Box-Cox-transformation ( $\lambda = 0,6$ ), fordi der var tydelig variansheterogenitet.

Parameter	Estimat	Standardfejl
<b>Bladlængde</b>		
b	0,708	0,074
d	70,818	1,361
ED50	105,415	13,970
<b>Strållængde</b>		
b	0,659	0,084
d	13,948	0,365
ED50	80,273	15,480
<b>Tørvægt</b>		
b	0,769	0,114
d	0,217	0,007
ED50	53,819	11,027
<b>Relativ vækstrate</b>		
b	0,888	0,132
d	0,152	0,003
ED50	116,046	18,365

Tabel 9.4 viser parameterestimer for de enkelte dosis-responsmodeller. Bladlængde og relativ vækstrate havde nogenlunde samme ED50, mens strållængde og tørvægt havde noget mindre ED50 værdier. Den følsomste responsvariable var tørvægt, som lå på ca. det halve af ED50 for bladlængde og vækstrate.



Figur 9.4. Estimerede dosis-responskurver for den tre-parameter Log-logistiske model anvendt på fire endpoints.

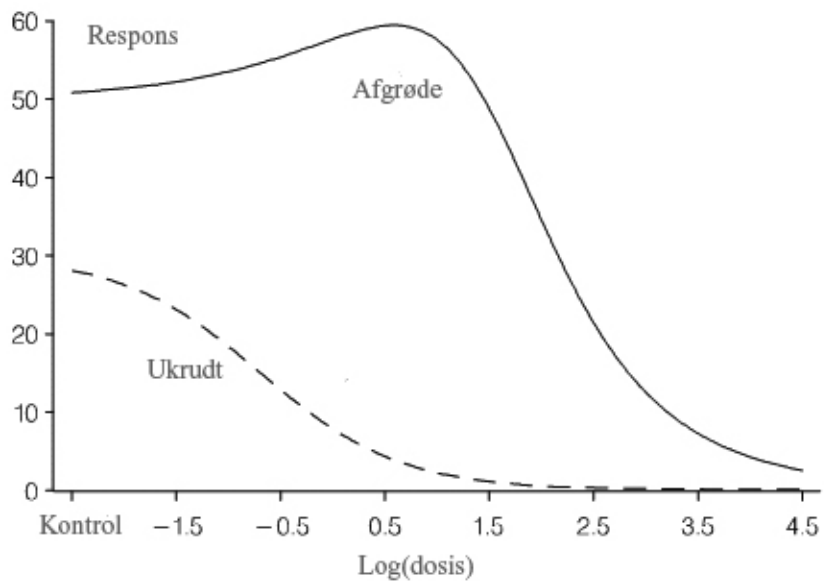
#### 9.2.4 Afsluttende bemærkninger

Fra et biologisk synspunkt viser især det sidste eksempel, hvor vigtigt det er at bruge biologisk viden til at reducere antallet af parametre for at få et toksicitetsestimater med en rimelig præcision.

### 9.3 Modificerede respons kurver

Forudsætningen for at benytte de dosis-responsmodeller, der er blevet brugt indtil nu, har været, at responsen enten er monotont aftagende eller stigende. I visse situationer kan man komme ud for, at denne forudsætning ikke er opfyldt (Figur 9.5). Der kan være flere grunde til dette, hvoraf nogle kan forklares ud fra samspillet mellem forskellige organismer, som påvirker hinanden i selve forsøget, eller det kan skyldes, at teststoffet ved meget lave doseringer påvirker organismen med andre virkningsmekanismer end den primære. Uanset årsagen kan man konstatere en stimulering, og ved meget små doseringer kan denne stimulering ikke umiddelbart forklares ud fra stoffets kendte virkning. Fænomenet har været kendt længe og kaldes hormesis eller et to-faset forløb, hvor kurverne har såvel en stigende som en faldende fase (Pestemer and Günther 1993; Cedergreen *et al.*, 2005).

Klassiske eksempler kan findes indenfor herbicider. Herbicider er fra producentens side designet til at bekæmpe uønsket plantevækst, og herbicidernes selektive brug i afgrøder er derfor et spørgsmål om dosis.

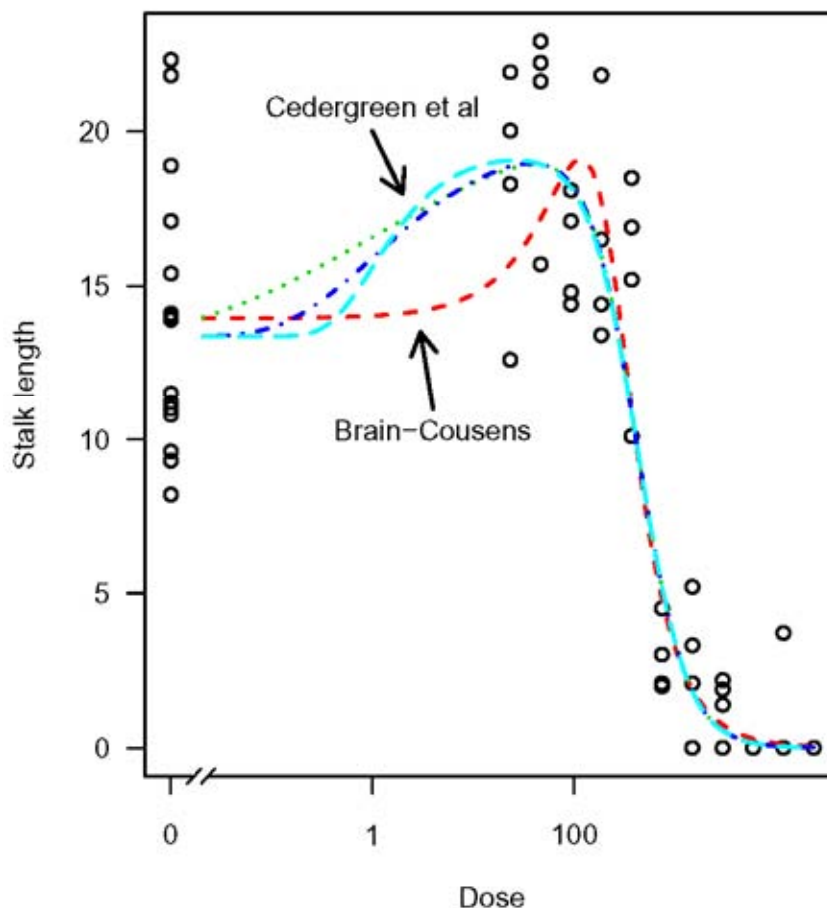


Figur 9.5. Dosis-respons for afgrøde og ukrudt som vokser sammen under påvirkning af voksende koncentrationer af et herbicid.

Stigningen i afgrøden i Figur 9.5 behøver ikke nødvendigvis at skyldes herbicidets virkning på afgrøden. Årsagen er snarere, at ukrudtet er mere følsomt overfor herbicidet end afgrøden er, hvorved konkurrencetrykket fra ukrudtet på afgrødeplanten mindskes med stigende dosis indtil også afgrødeplantens vækst påvirkes ved høje doseringer.

Mens sammenhængen i Figur 9.5 beskriver den ønskede virkning af et herbicid på afgrøde og ukrudt og derfor fra et bekæmpelsesmæssigt synspunkt er indlysende, så er der andre eksempler, der er knapt så lette at tolke. I et dosis-responsforsøg med regnorme fik man en tydelig hormesis-virkning ved stigende doser af et xenobiotika uden at kunne forklare det umiddelbart. Det viste sig efterfølgende, at en parasit i regnormen var mere følsom over for det pågældende xenobiotika end selve regnormen. Dermed fik regnormen en fordel ved små doser af det pågældende xenobiotika (Cedergreen 2007, personlig kommunikation). Et andet eksempel (med en binomialfordelt respons) stødte vi allerede på i afsnit 6.

Uanset om man har en forklaring på fænomenet eller ej vil en ikke-monoton faldende eller stigende dosis-responskurve kræve passende dosis-responssmodeller, der kan beskrive et sådant forløb. Der findes en række hormesis-modeller, som man kan bruge i disse tilfælde: Brain & Cousens, 1989; Schabenberger & Birch 2001 samt Cedergreen *et al.*, 2005. I Figur 9.6 har vi sammenlignet forskellige modeller til at "fange" hormesis. Det er tydeligt, at der er store forskelle på, hvorledes de enkelte kurver beskriver hormesis.



Figur 9.6. Illustration af hormesis for strållængde i byg ved doseringer af metsulfuron-methyl. Kurverne er forskellige modifikationer af den log-logistiske model (Brain-Cousens-modellen samt modellerne baseret på Ligning (2.4) og (2.5) med  $\alpha$  lig 0,25 og 0,50).

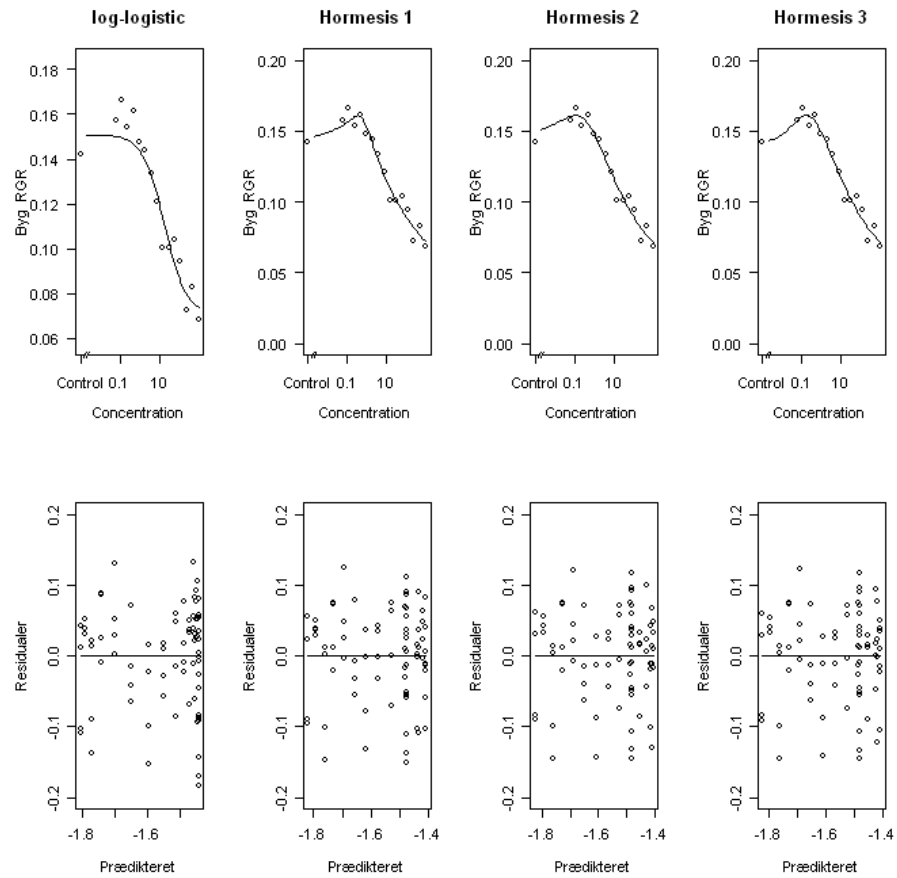
### 9.3.1 Datakilde 5

I et dosis-responsforsøg med byg og metsulfuron-methyl hvor det ikke var strållængden, men den relative vækstrate af byg, der var respons, observerede man også hormesis (Figur 9.7 og Tabel 9.5). Som det fremgår af Figur 9.5 (øverst til venstre) er den fire-parameter log-logistiske model ikke så god til at beskrive data. Det ses også i Tabel 9.5, hvor der er et signifikant lack-of-fit-test. Yderligere kan man i Figur 9.7 se, at det tilhørende residualplot (nederst til venstre) viser systematisk afvigelse ved lave doseringer: den log-logistiske kurve underestimerer responsen ved lave doseringer. Ved at betragte en af de 3 modeller, bliver lack-of-fit-testet stadig signifikant for de 2 første hormesis-modeller, men ikke for den sidste (Tabel 9.5). I Figur 9.7 ses det tydeligt på fordeling af residualer, at fordelingen er lidt bedre end f.eks. for den log-logistiske kurve. Forskellen på de tre hormesis-modeller er, at  $\alpha$  er fastholdt på tre forskellige værdier (nemlig 0,25; 0,50 og 1), som erfaringsmæssigt giver biologisk mening. Dette gøres, fordi der sjældent er data nok på den stigende del af kurven til at kunne estimere denne parameter statistisk.



Tabel 9,5, Sammendrag af anvendte dosis-responsmodeller for den relative vækstrate (RGR) med den log-logistiske model (Ligning (2.1)) og tre forskellige modifikationer af en hormesis model (Ligning (2.5) med  $\alpha$  lig 0,25, 0,50 og 1) for den relative vækstrate af byg ved stigende koncentrationer af metsulfuron-methyl. Box-Cox transformation blev benyttet ( $\lambda=0,3$ ).

Parameter	Estimat	Stand. Fejl
<i>Fire-parameter log-logistisk</i>		
b	0,797	0,179
c	0,071	0,007
d	0,151	0,003
ED50	18,651	8,382
Lack-of-fit-test: p= 0,0225		
<i>Hormesis med <math>\alpha=0,25</math></i>		
b	0,828	0,220
c	0,068	0,009
d	0,152	0,003
e	38,090	40,869
f	-0,016	0,019
Lack-of-fit-test: p = 0,0174		
<i>Hormesis med <math>\alpha=0,50</math></i>		
b	0,768	0,137
c	0,074	0,006
d	0,149	0,004
e	4,784	5,524
f	0,067	0,079
Lack-of-fit-test p= 0,0213		
<i>Hormesis med <math>\alpha=1</math></i>		
b	0,534	0,087
c	0,066	0,008
d	0,142	0,004
e	1,413	0,983
f	0,216	0,083
Lack-of-fit-test: p= 0,2521		



Figur 9.7. Estimerede dosis-responskurver for den relative vækstrate (RGR) og residualplot fra forskellige modeller i Tabel 2.1.

### 9.3.2 Afsluttende bemærkninger

Som det fremgår af ovenstående er det ikke altid, man kan forvente monotont stigende eller faldende dosis-responskurver. I sådanne tilfælde er det nødvendigt at bruge mere specialiserede dosis-responsmodeller, som kan tage hensyn til det to-fasede kurveforløb.

# 10 To-generationsreproduktionsforsøg

## 10.1 Problemstilling

Reproduktionsforsøg udføres for at undersøge, hvorvidt et stof er toksisk i relation til forplantning. Sædvanligvis gennemføres forsøget over to generationer, hvor stoffet hele tiden administreres til forældre/afkom i tre forskellige dosisgrupper og en kontrol. Under forsøgsperioden måles forskellige variable som f.eks. kuldstørrelse for 4 forskellige doseringer, hvoraf en er kontrolgruppen. Ud fra data fastlægges en No Observable Effect Level (NOEL), ED10 eller en Benchmark Dose (BMD). I litteraturen findes der flere definitioner af BMD, men her vil vi holde os til den givet i (Crump, 2002), som kort beskrives i det følgende. Vi betegner responsvariablen med  $Y$  og dosis-responsfunktionen med  $\mu(x)$ . Endelig lader vi  $P_0$  være sandsynligheden for uønsket respons ved dosis 0. Denne sandsynlighed sættes typisk til 5%. BMD er da den dosis  $x^*$ , der giver en forøget risiko svarende til en fastlagt Benchmark Risk (BMR) set i forhold til  $P_0$ :

$$P(x^*) - P_0 = BMR.$$

BMR sættes typisk til 5%. Dette fører til at BMD er den dosering  $x^*$ , der er fastlagt ud fra ligningen

$$\mu(x^*) = \mu(0) + \sigma(\Phi^{-1}(P_0) - \Phi^{-1}(P_0 + BMR))$$

hvor funktionen  $\mu$  er defineret ovenfor, mens  $\Phi^{-1}$  betegner den inverse funktion af fordelingsfunktionen for en standard normalfordeling. Den nedre 95%-konfidensgrænse for BMD benævnes BMDL, og denne angives typisk ved en risikovurdering.

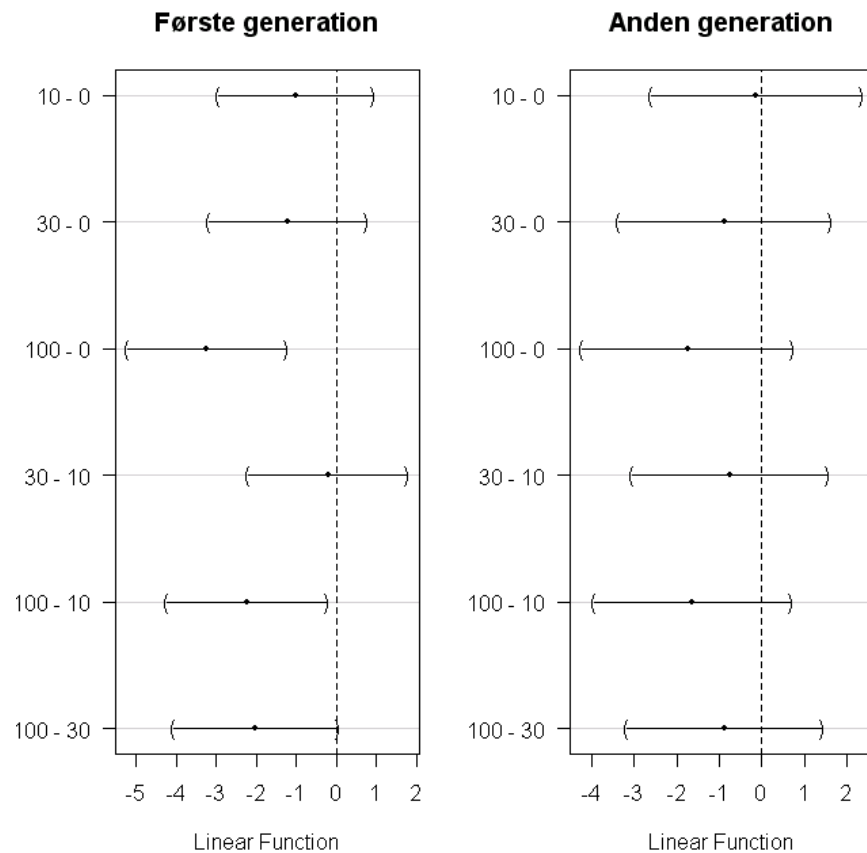
### 10.1.1 Kuldstørrelse

Her gives et eksempel med stoffet XL. Forsøget blev udført som et 2-generations forsøg med 26 kuld i hver af de 4 dosis grupper (0, 10, 30 og 100 mg/kg/dag). Forældredyr (første generation) blev fodret med stoffet i 10 uger før parring, under parringsperioden, drægtighed og laktation indtil aflivning. Anden generation blev eksponeret på tilsvarende måde. Der blev målt en række variable herunder kuldstørrelse (antal unger ved fødsel).

#### ***NOEL-tilgang***

I Figur 10.1 er de mest interessante forskelle dem mellem dosis 0 og de andre doseringer, og det fremgår tydeligt, at der er forskel mellem dosis 100 og 0 i første generation, mens der ikke er forskel mellem doseringer i anden generation. I en klassisk toksikologisk vurdering konkluderes der derfor, at kuldstørrelsen er signifikant reduceret i 100 mg/kg/dag gruppen i forhold til

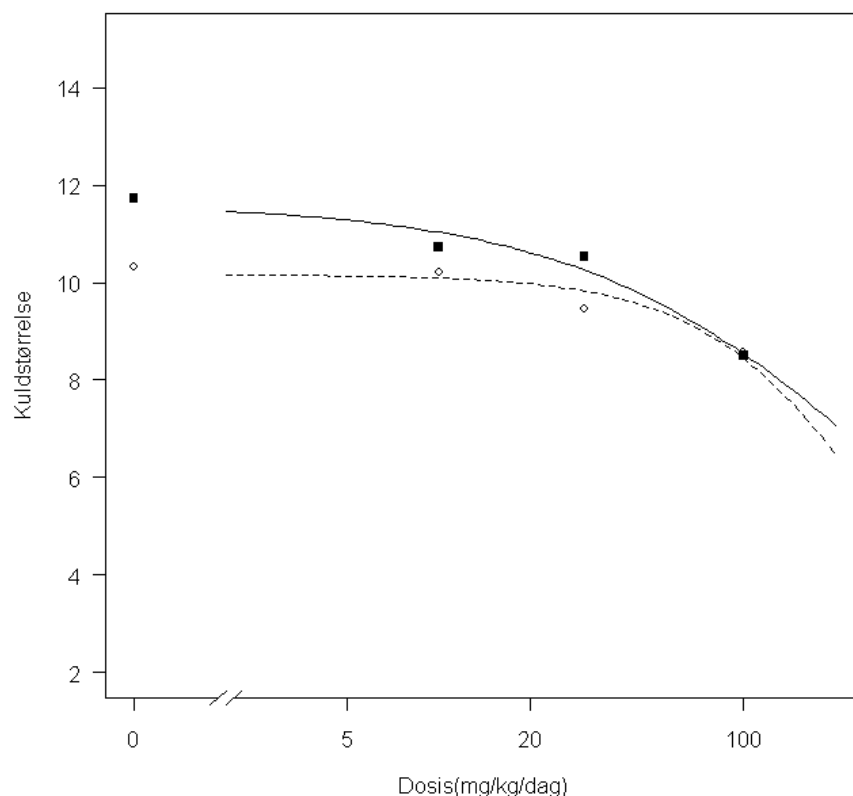
kontrol gruppen i første generation, men ikke i anden generation, mens 30 mg/kg/dag gruppen ikke er forskellig fra kontrolgruppe. NOEL vil derfor blive sat til 30 mg/kg/dag.



Figur 10.1. Resultat af en variansanalyse for hver af de to generationer. Grafen viser forskelle i kuld størrelse (x-akse) og tilhørende 95%-konfidensintervaller mellem de enkelte doseringer (parvise sammenligninger). Hvis intervallerne indeholder 0, er der ingen forskel på de tilhørende 2 doseringer.

### ***Dosis-responsanalyse***

Ved at benytte en tre-parameter log-logistisk model, hvor den logaritme-transformerede ED50 indgår som parameter (se afsnit 4.1), fik vi en øvre grænse på kuld størrelsen for første og anden generation på henholdsvis 12 (0,5) og 10 (0,7) dyr/kuld. Figur 10.2 viser, at de estimerede dosis-responsmodeller giver en ganske rimelig beskrivelse af data, og i begge tilfælde understøttes dette indtryk af ikke-signifikante lack-of-fit-tests.



Figur 10.2. Kuldstørrelse (dyr/kuld) ved fødsel som funktion af dosis af stoffet XL. Vi har brugt en tre-parameter log-logistisk model (Ligning (2.1b)). Første generation er den fulde optrukne linje og anden generation den stiplede linje.

Som det ses i Tabel. 10.1 er de estimerede ED-værdier for første generation omtrent en faktor 2 mindre end de tilsvarende estimer for anden generation. Ved at sammenligne ED50 i Tabel 10.1 med det faktiske doseringsområde, ses at de estimerede ED50 for begge generationer ligger uden for doseringsområdet, men det afspejler sig også direkte i konfidensintervallerne for ED50, som er meget brede.

Tabel 10.1. ED-estimer med tilhørende 95%-konfidensintervaller for de to generationer, baseret på dosis-responsmodellerne vist i Figur 10.1.

ED- niveau	(mg/kg/dag)	95%-konfidensinterval	
		Nedre	Øvre
Første generation			
10	23	4	130
20	64	27	151
50	368	75	1815
Anden generation			
10	42	5	381
20	120	31	473
50	732	20	26279

For både første og anden generation ligger de estimerede ED10-værdier ikke langt fra den estimerede NOEL på 30.

### ***Benchmark dose-proceduren***

Ved at bruge den ovenfor beskrevne metode for beregning af BMD fås, at BMD er 18 med tilhørende BMDL på 4 for første generation. For anden generation fås BMD til at være 70 med en tilhørende BMDL på 17. For

anden generation afviger BMDL-værdien noget fra nedre grænse i 95%-konfidensintervallet for ED10 baseret på dosis-responsanalysen, mens de tilsvarende værdier er ens for første generation.

#### 10.1.2 Afsluttende bemærkninger

Ovennævnte assays er særdeles bekostelige og derfor vil man økonomisere med de administrerede doser. Traditionen er at anvende variansanalyser og undersøge signifikante forskelle mellem effekter ved forskellige doseringer i relation til kontrolgruppen (NOEL). Ved brug af dosis-responsanalyser er det muligt at udregne ED10 og dermed få et mere brugbart estimat, der ikke nødvendigvis er betinget af de valgte doser.

# 11 Generel diskussion

Vi har i denne rapport set på symmetriske og asymmetriske samt ikke-monotone sigmoide dosis-responskurver med øvre og nedre asymptotiske grænser. For at kunne beskrive en dosis-respons kurves forløb og få præcist bestemte parametre er det vigtigt at få hele kurveforløbet bestemt på grundlag af data. De fleste doseringskurver er enten monotont stigende eller aftagende, selv om der godt kan optræde modeller som enten viser stimulering eller hæmning ved lave doseringer.

Vedrørende brugen af dosis-responsforsøg i toksikologisk og økotoksikologisk sammenhæng viser denne rapport, at der rent teknisk og statistisk er et stort sammenfald i metoder. Principperne er de samme for de to discipliner, men ofte vil der være forskelle i forsøgsplanerne. I toksikologien kan der være kraftige restriktioner på antallet af doseringer og antallet af dyr af etiske eller økonomiske årsager. Det medfører, at man ofte kun har tre til fire doseringer til rådighed, eller at man simpelthen tester en dosering ad gangen, indtil man finder en målbar effekt. Traditionelt har man valgt at prioritere gentagelser inden for samme dosering frem for at ind anvende flere doseringer med færre eller ingen gentagelser. Dette princip er sandsynligvis indført for at reducere den variation i testresultaterne som skyldes intra-species variation i følsomhed, da det er særdeles bekosteligt at fremavle homogene populationer af forsøgsdyr. Desuden har den statistiske analyse normalt begrænset sig til variansanalyse, og her er gentagelser en forudsætning for at kunne udføre den statistiske analyse til bestemmelse af NOEL/LOEL. Hvis man, som vi anbefaler (se diskussion nedenfor), forlader brugen af NOEL/LOEL som endpoints i toksikologiske og økotoksikologiske tests skal gænge forsøgsprotokoller tages op til revision under hensyntagen til begrænsninger i antallet af forsøgsdyr, pladsmangel mv.. En oplagt mulighed for at optimere forsøgsplaner i forhold til givne ressourcer er at anvende flere doseringer på bekostning af antallet af gentagelser. Vore simuleringer viser klart, at den højeste præcision af relevante parameterestimer, f.eks. ED10 og ED50, fås ved at bruge flere doseringer på bekostning af gentagelser. I denne forbindelse har vi også behandlet problemet med udvælgelse af hvilke doseringer man bør benytte for at få den mest præcise beskrivelse af data. Valget må i alle tilfælde bygge på tidligere erfaringer med de pågældende pesticider og dermed en formodning om, hvorledes forløbet forventes at være af dosis-responskurven er. Dette understreger vigtigheden af at der udføres indledende screeningsforsøg til fastlæggelse af doseringer i de definitive test. Ved screeningsforsøg kan man igen drage fordel af at reducere antallet af gentagelser for at kunne dække et større spænd af doseringer. Efter analyser af screeningsforsøg kan den endelige doseringsrække fastlægges, og det skal nævnes at i langt de fleste af de forsøg vi har gennemgået har doseringerne være jævnt fordelt på en logaritmisk dosisakse i overensstemmelse med almindelig praksis ved design af toksikologiske og økotoksikologiske test (Finney, 1978).

Som nævnt gælder principperne for øget antal doseringer på bekostning af antallet af gentagelse kun, når analysen er baseret på dosis-responsanalyser med asymptotiske øvre nedre grænser. Hvis man ønsker at finde en

NOEC/NOED eller LOEC/NOED (principperne herfor er ikke behandlet i nærværende rapport), gælder det, som ved alle parvise sammenligninger, om at have så mange gentagelser så muligt indenfor de enkelte doseringer. Kun herved er det muligt at detektere små forskelle i effekter med en variansanalyse. Der i dag er en righoldig litteratur, som dokumenterer fordelene ved brug af ikke-lineære analyser af toksicitetsdata frem for NOEC/LOEC (bl.a. Chapman and Caldwell 1996; Kooijman 1996), og ud fra et videnskabeligt synspunkt burde emnet dermed være uddebatteret. I den administrative anvendelse af toksikologiske og økotoksikologiske data ser man dog stadig at NOEC/LOEC anbefales som endpoints. Dette er f.eks. tilfældet selv i den nyeste revision af pesticid-lovgivningen 91/414 og dette altså på trods af den videnskabelige verdens afvisning af disse parametres brugbarhed. Der er således en stor træghed i at kuldaste knæsatte principper i den administrative anvendelse af testresultaterne og dette har naturligvis en sammenhæng med at praktiske (f.eks. den store mængde af data der allerede findes, og som har fokuseret på NOEC/LOEC) såvel som politiske hensyn spiller en stor rolle forbindelse med udarbejdelser af standarder og guidelines. Det er dog projektgruppens forhåbning, at de øgede muligheder for statistisk fortolkning af ikke-lineære dosis-respons vil bidrage til at forsøgsplaner baseret på bestemmelse af NOEC/LOEC helt bortfalder i de internationale standarder og vejledninger.

Overordnet set er afrapporteringen af et dosis-responsforsøg den samme inden for toksikologi og økotoksikologi, og den bør som minimum indeholde dokumentation for, om den valgte model til beskrivelse af data reelt beskriver variationen i data på en tilfredsstillende måde. Når det drejer sig om normalfordelte data med gentagelser inden for doseringer kan en sammenligning mellem dosis-responsmodellen med en almindelig variansanalysemodel, give et fingerpeg om regressionsmodellen beskriver data tilfredsstillende. I de tilfælde, hvor man ikke har gentagelser, kan man som vist i rapporten benytte sig af en test baseret på kumulerede residualer. Selvom det kan synes trivelt, skal det dog alligevel understreges at en visuel bedømmelse altid skal indgå i en vurdering af, om en given model beskriver data tilfredsstillende.

De udførte analyser i denne rapport viser at der til bestemmelse af 50%'s effekt-værdier (LC50/LD50/EC50/ED50) ikke vil være stor forskel på, om man benytter en symmetrisk eller asymmetrisk kurve til at beskrive dosis-responsforløbet. Er man derimod interesseret i at bestemme de koncentrationer, hvor den toksiske virkning begynder at indtræde (ofte defineret som 10%'s effekt – LC10/LD10/EC10/ED10) kan der være betydelige forskelle afhængigt af modelvalget. Ofte ses det inden for såvel toksikologi som økotoksikologi, at valget mellem en symmetrisk model (f.eks. log-logistisk) eller asymmetrisk model (f.eks. Weibull) er et spørgsmål om tradition. Hvis man har veldefinerede data med lille varians, viser de udførte analyser, at en asymmetrisk model bør afprøves da denne måske bedre vil kunne beskrive data end en symmetrisk. Med udgangspunkt i de tilvejebragte problemstillinger under dette projekt, er det dog vores erfaring, at variationen i data ofte vil være så stor det vil være af mindre betydning for estimationen af effektværdier om man vælger en symmetrisk eller asymmetrisk model. Dette er fundet uanset om høje eller lave effektiveauer skal estimeres.

Sammenligninger af dosis-responskurver, som er udført uafhængigt af hinanden enten i tid eller rum, udgør en særlig udfordring i forbindelse med den statistiske databehandling. Her viser vore analyser, at man med fordel kan



benytte blandede ikke-lineære modeller, hvor der justeres for variationen fra kurve til kurve og dermed opnå et fælles estimat for en ED<sub>x</sub>/EC<sub>x</sub> værdi. Sådanne modeller kan også bruges til at analysere dosis-responskurver med gentagne målinger på samme individ. En rutinemæssig brug af disse blandede ikke-lineære modeller er desværre ikke tilgængelig i øjeblikket, fordi der ofte vil være problemer med konvergens i estimationen af den ikke-lineære regression.

For data opnået i forsøg med kontinuerte respons er det vigtigt, at der ikke er variansheterogenitet i datasættet. Normalt vil der være variansheterogenitet, hvis forskellen mellem maksimal og minimum respons er stor, f.eks. hvis maksimum er 4-5 gange større end minimum respons. Derfor skal man teste for variansheterogenitet ved statistisk analyse af kontinuerte data og tage højde for denne heterogenitet, hvis den kan påvises. I dette projekt har vi benyttet en såkaldt Box-Cox transformation af såvel højre som venstre side af regressionsmodellen. Erfaringen viser, at parameterestimererne, herunder EC<sub>50</sub> og EC<sub>10</sub>, ikke påvirkes synderlig af, om der er variansheterogenitet eller ej. Derimod vil estimationen af standardfejlen på parametrene blive påvirket. Da vurderinger af et stofs toksicitet og/eller økotoksicitet netop er en kombination af parameter estimatet (f.eks. EC<sub>50</sub>/EC<sub>10</sub>) og den tilhørende usikkerhed, kan det få vidtrækkende følge, hvis f.eks. den nedre 95% konfidensgrænse ikke er bestemt korrekt pga. at man ikke har taget højde for variansheterogenitet. Går man videre med den fejlagtigt estimeret værdi ved f.eks. fastsættelse af en grænseværdi er det dermed givet, at denne vil blive baseret på værdier, der ud fra et statistisk synspunkt ikke er holdbare.

Ved sammenligning af dosis-respons kurver støder man af og til på det problem, at de fundne kurver har forskellige øvre og/eller nedre grænser. Disse grænser spiller en stor rolle for bestemmelsen af LC<sub>50</sub>/LD<sub>50</sub>/EC<sub>50</sub>/ED<sub>50</sub>-værdien, da denne er defineret som doseringen midt mellem den øvre og den nedre grænse. For binomialfordelte data, som har et responsområde mellem 0 og 1, er det også muligt at have forskellige øvre og nedre grænser, men den nedre grænse kan aldrig blive være negativ. For kontinuerte data er forholdene yderligere kompliceret af, at såvel den øvre som den nedre grænse ikke har "naturlige" værdier. Altså kan der for kontinuerte data godt forekomme negative responsværdier (dvs. at den nedre grænse forrykkes), som det er tilfældet, hvis man undersøger vækstræthæmning i en algetest og cellerne dør i løbet af testperioden. Desuden er det efterhånden veldokumenteret at en lang række pesticider og andre kemiske stoffer virker stimulerende ved meget små doser, og dermed fås også en afvigelse fra det forventede monotone kurveforløb - et fænomen der kaldes hormesis. Her er den øvre grænse for ubehandlede organismer mindre end for organismer, der har fået meget små doseringer. I begge de beskrevne tilfælde findes der modeller til at tage højde for disse fænomener, men i toksikologisk og økotoksikologisk sammenhæng besværliggøres målinger af et givet stofs toksicitet og der vil traditionelt være forskellige opfattelser af hvad man skal gøre for at få en tilfredsstillende beskrivelse af dosis-respons forløbet. Der pågår en videnskabelig debat af disse problemstillinger, og det har ligget uden for dette projekts rammer at fastlægge en "best practice" på dette område. Dog giver de gennemgåede eksempler forslag til hvordan man kan håndtere denne problemstilling.

For binomialfordelte data er det en udbredt praksis, at data analyseres som procentvis virkning, f.eks. procent døde eller immobile dyr, og at man herefter behandler data som om de er kontinuerte. Simuleringsstudierne udført i denne rapport viste, at der ikke er den store forskel mellem parameterestimerer, f.eks.

LD50, når data analyseres korrekt dvs. med binomialfordelingen, eller når man benytter procentvis virkning og betragter data som normalfordelte (eller evt. logaritmisk normalfordelte). Til gengæld har det stor betydning for estimationen af standardfejlen om man benytter den "forkerte" model til estimation af dosis-respons kurven. Dermed er situationen altså tilsvarende den vi har beskrevet ovenfor for normalfordelte data, altså at f.eks. det nedre 95% konfidensinterval ikke vil blive bestemt korrekt med risiko for fejlagtige administrative følgevirkninger.

Da fortolkning af resultater fra toksikologiske og økotoxikologiske undersøgelser kan få vidtrækkende betydning i reguleringsmæssig sammenhæng (de danner f.eks. udgangspunkt for afskæringsværdier i registrering og klassifikation i forskellige fareklasser/ kategorier) er det bydende nødvendigt, at resultaterne afrapporteres på den videnskabeligt set mest korrekte måde. Således vil estimeringen af præcise konfidensintervaller være af stor betydning. Dette gælder naturligvis i de tilfælde hvor en administrativ beslutning skal træffes på et sparsomt datagrundlag, men det er vigtigt at understrege at det ligeledes gør sig gældende i de tilfælde, hvor man på baggrund af store datamængder antager, at man har testet de mest følsomme arter. I dette tilfælde vil man normalt reducere sikkerhedsfaktorerne ved fastsættelse af f.eks. vandkvalitetskriterier til et minimum (evt. slet ikke anvende sikkerhedsfaktorer). Dette er f.eks. blevet foreslået for akvatiske makrofytter i de tilfælde, hvor man har fulde arts-følsomhedskurver (Species Sensitivity Distributions, SSDs). Sammenligninger af SSD'er for mikro- og makrokosmosforsøg har vist, at den nedre 95%-konfidensgrænse for 5% farlighedskoncentrationen ( $HC_5$ ) kan antages at være beskyttende for den akvatiske flora (Van den Brink *et al.*, 2006), men med en forventet øget brug af SSD'er i reguleringsmæssig sammenhæng i fremtiden, vil en korrekt bestemmelse af konfidensintervaller på ikke-lineære regressioner, som foreslået i denne rapport få en øget betydning. I de tilfælde, hvor man pga. af datakvalitet og/eller -mængde anvender meget store sikkerhedsfaktorer i risikovurderingssammenhæng anbefales vi på baggrund af dette projekt at man tager udgangspunkt i LC50/LD50/EC50/ED50-værdier, som altid vil være bestemt med mindre usikkerhed end LC10/LD10/EC10/ED10-værdier. Ved risikovurderinger, hvor sikkerhedsfaktorer i størrelsesordenen 10-1000 anvendes vil en usikkerhed på konfidensgrænserne på under en faktor to måske umiddelbart kunne synes kun at være af akademisk betydning. I praksis vil de nemlig sjældent vil rykke ret meget ved den endelige risikovurdering. Imidlertid bør man ikke glemme, at en korrekt statistisk analyse af data sikrer en optimal udnyttelse af de tilvejebragte data og beforder videnskabeligt baserede vurderinger af datakvaliteten. Hermed styrkes også mulighederne for en troværdig risikokommunikation mellem myndigheder, virksomheder og andre interessenter.

# 12 Konklusion

Projektets analyser af problemstillinger knyttet til kvantificering af toksikologiske og økotoksikologiske dosis-respons forsøg har ført til følgende konklusioner:

- Der opnås større præcision på estimater af effekt-koncentrationer/-doseringer ved at øge antallet af doseringer frem for at øge antal replikater.
- Modelvalg og -kontrol understøttes bedre af et testdesign med mange doseringer og få replikater end af et design med få doseringer og mange replikater.
- Modelvalg påvirker i høj grad usikkerhedsbestemmelsen på estimater af effekt-koncentrationer/-doseringer.
- Modelvalg har ikke stor indflydelse på estimatet for LC50/LD50/EC50/ED50.
- Transformationer kan afhjælpe afvigelser fra antagelsen om normalfordeling og fjerne variansheterogenitet.
- Binære data bør analyseres ved brug af binomialfordelingsmodeller idet usikkerhedsbestemmelsen ellers ikke vil være korrekt

Der tegner sig således et billede af, at regressionsbaserede dosis-responsmodeller er i stand til langt mere fleksibelt at afspejle og beskrive strukturerne i data fra både human- og økotoksikologiske dosis-responsforsøg, end tilfældet er for variansanalysebaserede fortolkninger af dosis-respons forsøg.

Da fortolkning af toksikologiske og økotoksikologiske data kan få vidtrækkende konsekvenser for f.eks. pesticid registrering og fastsættelse af vandkvalitetskriterier understreger rapportens anbefalinger at en korrekt statistisk analyse er nødvendig for at sikre en optimal udnyttelse af de tilvejebragte data.

Det gælder også selvom det set fra et pragmatisk synspunkt måske kun medfører mindre ændringer i de estimerede effektværdier, da en troværdig risikokommunikation mellem myndigheder, virksomheder og andre interessenter må baseres på, at de data man administrativt benytter i risikovurderinger er fremkommet i overensstemmelse med den bedste praksis set fra et videnskabeligt synspunkt.



# 13 Perspektivering

Som vist i dette projekt findes der i dag egnede (og veletablerede) statistiske metoder til analyse af resultater opnået i dosis-respons baserede toksikologiske og økotoksikologiske bioassays. På baggrund af det gennemførte projekt opfordrer vi til, at de regulerende myndigheder i internationale fora tager initiativ til udfasning af utilstrækkelige, forældede og endda til tider ukorrekte metoder og koncepter for databehandling. Hermed vil man kunne opnå en langt bedre udnyttelse af såvel eksisterende som nye toksikologiske/økotoksikologiske resultater. I særlig grad opfordres til at de fastlåste anbefalinger af antal gentagne målinger per dosering og LOEL/NOEL-principper til bestemmelse af "nul-effekt" niveauer udfases. I stedet anbefaler vi, at man opstiller krav om anvendelse af korrekte statistiske metoder, der tilgodeser datas beskaffenhed. Her er det vores håbe, at projektets konklusioner kunne bidrage, og vi mener, at der i en administrativ sammenhæng i særlig grad er behov for:

- promovning af regressionsbaserede dosis-responsmodeller i risikovurdering i bred forstand,
- forskydning af fokus fra ANOVA-præget punkttestimation (NOEC/ED) og over mod usikkerhedsbestemmelse af punkttestimater (benchmark dose-begrebet),
- paradigmeskift over mod - i hvert fald delvist - statistik-baserede usikkerhedsfaktorer i hele risikovurderingsprocessen,
- afklaring og afrunding af begreber, definitioner og fortolkninger af regressionsbaserede dosis-responsmodeller (bl.a. benchmark dose-begrebet).

På den forskningsmæssige side finder vi på baggrund af det udførte projekt, at der er behov for statistisk metodeudvikling inden for følgende emner:

- afklaring af definitioner og metodeudvikling i forbindelse med benchmark dose-metodologien,
- meta-analyse af resultater fra forskellige forsøg med samme biologisk aktive stof,
- dosis-responsmodellering af flerdimensionale endpoints samt ordinale data (dvs. data som kan rangordnes såsom subjektive scores som eksempelvis: lav, middel og høj) og tælldata.



# Litteratur

- Baun, A., Andersen, J. S. & Nyholm, N. (1999). Correcting for toxic inhibition in quantification of genotoxic response in the umuC test, *Mutation Research-Genetic Toxicology and Environmental Mutagenesis* 441, 171-180.
- Brain, P. & Cousens, R. (1989). An equation to describe dose responses where there is stimulation of growth at low doses, *Weed Research* 29, 93-96.
- Cabanne, F., Gaudry, J. C. & Streibig, J. C. (1999). Influence of alkyl oleates on efficacy of phenmedipham applied as an acetone:water solution on *Galium aparine*, *Weed Research* 39, 57-67.
- Carroll, R. J. (2003). Variances Are Not Always Nuisance Parameters, *Biometrics*, 59, 211-220.
- Cedergreen, N, & Ritz, C, & Streibig, J, C. (2005). Improved empirical models describing hormesis, *Environmental Toxicology and Chemistry* 24, 3166-3172.
- Cedergreen, N., Kudsk, P., Matthiasen, S. & Streibig, J. C. (2007). Combination effects of herbicides: Do species and test system matter?, *Pest Management Science* 63, 282-295.
- Cedergreen, N. (2008). Herbicides can stimulate plant growth, *Weed Research* 48, (In press).
- Chapman, P. M. and Caldwell, R. S. (1996). A warning: NOECs are inappropriate for regulatory use, *Environmental toxicology and Chemistry* 5, 77-79.
- Christensen, M. G., Teicher, H. B. and Streibig, J. C. (2003). Linking fluorescence induction curve and biomass in herbicide screening, *Pest Management Science* 59, 1303-1310.
- Christensen, A. M., Ingerslev, F. & Baun, A. (2006). Ecotoxicity of mixtures of antibiotics used in aquacultures. *Environmental Toxicology and Chemistry* 25, 2208-2215.
- Cleveland, W, S. (1979). Robust locally weighted regression and smoothing scatterplots, *J. Am. Statist. Assoc.* 74, 829-836.
- Crump, K. (2002). Critical issues in benchmark calculations from continuous data. *Critical Reviews in Toxicology* 32, 133-153.
- Dette, H., & Neymeyer, N., & Pilz, K., F. (2005). A note on Non-parametric Estimation of the Effective Dose in Quantal Bioassay, *J. Am. Stat. Assoc.* 100, 503-510,

- Draper, N. R. and Smith, H. (1981). *Applied Regression*. John Wiley & Sons, Toronto.
- Environment Canada (2005). *Guidance Document on Statistical Methods for Environmental Toxicity Tests*, Method Development and Applications Section, Environmental Technology Centre, Environment Canada, Ottawa.
- Finney, D. J. (1965). The meaning of bioassay, *Biometrics* 21, 785-810.
- Finney, D. J. (1978). *Statistical Method in Biological Assay*, 2 ed, Charles Griffin & Company Ltd, London.
- Finney, D. J. (1979). Bioassay and the practice of statistical inference, *International Statistical Review* 47, 1-12.
- Hass U., Scholze M., Christiansen S., Dalgaard M., Vinggaard A. M., Axelstad M., Broeng Metzdorff S. & Kortenkamp A. (2007). A Combined exposure to anti-androgens exacerbates disruption of sexual differentiation in the rat, *Environmental Health Perspectives* 115, 122-8.
- Hewlett, P. S. and Plackett, R. L. (1979). An introduction to the interpretation of quantal responses in Biology, 1 ed, Edward Arnold, London.
- ISO 13829. *Water quality - Determination of the genotoxicity of water and wastewater using the "umu-test"*, International Organization of Standardization, Geneva.
- ISO 15522. *Water quality - Determination of the inhibitory effect of water constituents on the growth of activated sludge microorganisms*, International Organization of Standardization, Geneva.
- ISO 6341. *Water quality - Determination of the inhibition of the mobility of Daphnia magna Stratus (Cladocera, Crustacea)*, International Organization for Standardization, Geneva.
- ISO 8692. *Water quality - Freshwater algal growth inhibition test with unicellular green algae*, International Organization of Standardization, Geneva.
- Jerne, N. K. & Wood, E. C. (1949). The validity and meaning of the results of biological assays, *Biometrics* 5, 273-299.
- Kooijman, S.A.L.M. (1996). An alternative for NOEC exists, but the standard model has to be abandoned first, *Oikos* 75, 310-316.
- Laskowski, R. (1995). Some good reasons to ban the use of NOEC, LOEC, and related concepts in ecotoxicology, *Oikos* 73, 140-144.
- Lin, D. Y. & Wei, L. J. & Ying, Z. (1993). Checking the Cox model with cumulative sums of martingale-based residuals, *Biometrika* 80, 557-572.
- Lin, D. Y. & Wei, L. J. & Ying, Z. (2002). Model-Checking Techniques Based on Cumulative Residuals, *Biometrics* 58, 1-12,



- Loader, C. (1999). *Local Regression and Likelihood*, Springer, New York.
- McCullagh, P. & Nelder, J. A. (1989). ***Generalized Linear Models***, Chapman & Hall, Boca Raton.
- Nottingham, Q. J. & Birch, J. B. 1996, Model-Robust Quantal Regression, Technical report 96-6, Virginia Polytechnic Institute and State University, Blacksburg.
- Nottingham, Q. J. & Birch, J. B. (2000). A semiparametric approach to analysing dose-response data, ***Statist. Med.*** 19, 389-404,
- Nyholm, N. (1985). Response variable in algal growth inhibition tests- biomass or growth rate? ***Water Research*** 19, 273-279.
- Nyholm, N. (1990). Expression of results from growth inhibition toxicity tests with algae, ***Archives of Environmental Contamination and Toxicology*** 19 518-522.
- Nyholm, N. 1994, Comments on response parameter in algal toxicity tests. ISO.
- OECD 1998, Report of the OECD Workshop on Statistical Analysis of Aquatic Toxicity data, ***OECD Environmental Health and Safety Publications, Series on Testing and Assessment, Paris***
- OECD 2003, Draft Guidance Document on the Statistical Analysis of Ecotoxicity Data, ***OECD Environmental Health and Safety Publications, Series on Testing and Assessment, Paris.***
- Park, D, & Park, S. (2006). Parametric and nonparametric estimators of  $EC_{100\alpha}$ , ***J. Stat. Comput. Simul.*** 76, 661-672.
- Pestemer, W. and Günther, P. (1993). No-Observable-Effect level (NOEL). I ***Herbicide Bioassays*** (J. C. Streibig and P. Kudsk (eds)), CRC Press, Boca Raton.
- Pinheiro, J. C. & Bates, D. M. (2000). *Mixed-Effects Models in S and S-PLUS*, Springer, New York.
- Ritz, C. & Streibig, J. C. (2005). Bioassay analysis using R. ***Journal of Statistical Software*** 12, 1-22.
- Ritz, C., Cedergreen, N., Jensen, J. E. & Streibig, J. C. (2006). Relative potency in nonsimilar dose-response curves. ***Weed Science*** 54, 407-412.
- Ritz, C. & Martinussen, T. (2008). The use of lack-of-fit tests in dose-response analysis. ***Submitted to Environmetrics.***
- Sand, S., von Rosen, D., Victorin, K. & Falk Phillipsson, A. (2006). Identification of a Critical Dose Level for Risk Assessment: Developments in Benchmark Dose Analysis of Continuous Endpoints, ***Toxicological Sciences*** 90, 241-251.

- Schabenberger, O. & Birch, J. B. (2001). Statistical dose-response models with hormetic effects. *Human and Ecological Risk Assessment* 7, 891-908.
- Seber, G. A. F., & Wild, C. J. (1989). *Nonlinear Regression*, Wiley & Sons, New York.
- Soto, A. M., Sonnenschein, C., Chung, K. L., Fernandez, M. F., Olea, N. & Serrano, F. O. (1995). The E-Screen Assay As A Tool to Identify Estrogens - An Update on Estrogenic Environmental-Pollutants. *Environmental Health Perspectives* 103, 113-122.
- Streibig, J. C. (1980). Models for curve-fitting herbicide dose response data. *Acta Agriculturae Scandinavica* 30, 59-64.
- Streibig, J. C. (1983). Joint action of root-absorbed mixtures of DPX-4189 and linuron in *Sinapis alba* L. and barley. *Weed Research* 23, 3-9.
- Streibig, J. C. (1984). Measurement of phytotoxicity of commercial and unformulated soil- applied herbicides. *Weed Research* 24, 327-331.
- Streibig, J. C. (1988). Herbicide bioassay. *Weed Research* 28, 479-484.
- Streibig, J. C., Rudemo, M. & Jensen, J. E. (1993). Dose-response curves and statistical models. I *Herbicide Bioassays* (J. C. Streibig and P. Kudsk (eds)), CRC Press, Boca Raton.
- Su, J. Q. & Wei, L. J. (1991). A Lack-of-Fit Test for the Mean Function in A Generalized Linear Model, *J. Am. Statist. Assoc.* 86, 420-426,
- Van Leeuwen C. J. & Hermens J. L. M. (eds) (1995). Risk Assessment of Chemicals: An Introduction, Kluwer Academic Publishers, Dordrecht.
- Vinggaard A. M. (2003). The combined effects of vinclozolin and procymidone do not deviate from expected additivity in vitro and in vivo, *Toxicological Sciences* 71, 251-262.
- Vølund, A. (1978). Application of the four-parameter logistic model to bioassay: comparison with slope ratio and parallel line models. *Biometrics* 34, 357-365.
- Vølund, A. (1980). Multivariate bioassay. *Biometrics* 36, 225-236.
- Vølund, A. (1982). Combination of multivariate bioassay results. *Biometrics* 38, 181-190.
- Wheeler, M, W, & Park, R, M, & Bailer, A, J. (2006). Comparing median lethal concentration values using confidence interval overlap or ratio tests, *Environmental Toxicology and Chemistry* 25, 1441-1444,
- Wright, S, E, & Bailer, A, J. (2006). Optimal Experimental Design for a Nonlinear Response in Environmental Toxicology, *Biometrics* 62, 886-892,

## Regneeksempel med drc

Rød tekst er R-kode, blå tekst er R-resultat og sort tekst er kommentarer.  
(Data fra afsnit 5)

```

library(drc)#Load the add on package to R
head (cabanne)# udskriv de første 6 observationer.
Dataset #indeholder 2 kurver(curve)
  dose drymat curve
1    0   1146    0
2    0   1005    0
3    0    756    0
4    0   1108    0
5    0    956    0
6    0    989    0

m1 <- drm(drymat~dose,curve, data=cabanne, fct=LL.4())
# Tilpas den 4 parameter log-logistic kurve LL.4

Control measurements detected for level: 0

plot(fitted(m1),residuals(m1),ylim=c(-400,400))
abline(h=0)
# Laver et residualplot af regressionen m1 med en linje
#gennem 0.

m2 <- drm(drymat~dose,curve,
data=cabanne, fct=LL.4(),adjust="bc1")
# Da fittet m1 viser kraftig variansheterogenitet køres
#analysen igen men nu med en Box-Cox transformation
#(adjust="bc1")

Control measurements detected for level: 0

plot(fitted(m2),residuals(m2),ylim=c(-2,2))
abline(h=0)

# Laver et residualplot af regressionen m2 med en linje
#gennem 0. Nu er variansheterogeniteten væk

anova(m2) #Laver et test for lack of fit, Det er ikke
#signifikant, altså er regressionsmodellen god nok
Lack-of-fit test

```

	ModelDf	RSS	Df	F value	p value
Two-way ANOVA	219	67.203			
DRC model	231	72.228	12	1.3647	0.1845

```

plot(m2) #Tegner en graf af data (gennemsnit pr
doser) #og regressionslinjer
summary(m2)# viser et sammendrag af
regressionsparametre

Model fitted: LL.4()

Parameter estimates:

      Estimate Std. Error   t-value   p-value
b:1    1.58975    0.39935    3.98079    1e-04
b:2    1.80114    0.18006   10.00276   4.197e-20
c:1   500.71562   18.65118   26.84632   2.868e-73
c:2   156.07046    8.80688   17.72142   2.992e-45
d:0   941.20777   21.15757   44.48563   1.455e-115
d:1   981.08700   36.55833   26.83621   3.063e-73
d:2  1016.21177   28.90036   35.16260   5.200e-95
e:1    51.04937    9.39667    5.43271   1.405e-07
e:2    86.53578    6.88103   12.57600   2.814e-28

Residual standard error: 0.5591745 (231 degrees of
freedom)

Non-normality/heterogeneity adjustment through optimal
Box-Cox transformation

Estimated lambda: 0.2
Confidence interval for lambda: [0.0110,0.353]

m3 <- drm(drymat~dose,curve,data=cabanne,
fct=LL.4(),pmodels=data.frame(curve,curve,1,curve),adju
st="bc1")
# m3 forudsætter at den over grænse er ens for de to
#kurver. Dette gøres med argumentet:
#pmodels=data.frame(curve,curve,1,curve)

Control measurements detected for level: 0
anova(m3,m2)# Test for lack of fit mellem regression
med forskellig forskellige parametre for de to kurver
og regression med fælles øvre grænse for de to kurver

1st model
  fct:      LL.4()
2nd model
  fct:      LL.4()

ANOVA table

      ModelDf    RSS   Df F value p value
1st model     233 73.702
2nd model     231 72.228   2  2.3565  0.0970

plot(m3) # Regressionsgraf

```

```

summary(m3) #Sammendrag for parametreestimerer

Model fitted: LL.4()

Parameter estimates:

              Estimate Std. Error  t-value  p-value
b:1           1.64480    0.33772   4.87023 2.058e-06
b:2           1.96297    0.17567  11.17432 8.524e-24
c:1          501.71049   17.82464  28.14701 3.216e-77
c:2          159.12111    8.32730  19.10838 6.653e-50
d:(Intercept) 972.59835   14.99596  64.85737 2.047e-151
e:1           52.55449    7.60871   6.90715 4.669e-11
e:2           94.14810    5.73829  16.40700 4.971e-41

Residual standard error: 0.5624205 (233 degrees of
freedom)

Non-normality/heterogeneity adjustment through optimal
Box-Cox transformation

Estimated lambda: 0.2
Confidence interval for lambda: [0.0110,0.353]

ED(m3,c(10,50),ci="delta")# beregning af ED10 og ED50
med konfidensintervaller

Estimated effect doses
(Delta method-based confidence interval(s))

      Estimate Std. Error  Lower  Upper
1:10  13.8182    4.3170  5.3129  22.323
1:50  52.5545    7.6087 37.5638  67.545
2:10  30.7391    4.2870 22.2929  39.185
2:50  94.1481    5.7383 82.8425 105.454

ED(m3,973*c(.9,.8,.5),ci="delta", type="absolut")
# Beregning af ED ud fra absolutte værdier i forhold
til den estimerede fælles øvre grænse.Se blandt andet
afsnit 9

Estimated effect doses
(Delta method-based confidence interval(s))

      Estimate Std. Error  Lower  Upper
1:875.7  23.1209    5.1829 12.9096  33.332
1:778.4  42.3775    6.4464 29.6768  55.078
1:486.5    NA         NA      NA      NA
#eksisterer ikke for curve nr 1 derfor NA=Not
Applicable
2:875.7  33.9737    4.4591 25.1884  42.759
2:778.4  52.1484    5.0879 42.1243  62.172
2:486.5 115.1509    6.3217 102.6959 127.606
Warning message:
In log((100 - p)/100) : NaNs produced

```

```

#Sammendrag af grafik
par(mfrow=c(2,2))#giver 4 grafer pr plot
plot(fitted(m1),residuals(m1),ylim=c(-400,400),
ylab="Ustandardiserede residualer",xlab="Prædikterede værdier",main="Variansheterogenitet")
abline(h=0)
plot(fitted(m2),residuals(m2),ylim=c(-2,2),
ylab="Ustandardiserede residualer",xlab="Prædikterede værdier", main="Box-Cox transformation")
abline(h=0)
plot(m2,main="Hver Kurve sine egne parametre")
plot(m3,main="Fælles øvre grænseparametre")

```

