

Berichte
aus der Biologischen Bundesanstalt für Land- und Forstwirtschaft

Reports

from the Federal Biological Research Centre for Agriculture and Forestry

Heft 17

1996

Fachgespräch zur Statistik in der Ökotoxikologie

26.- 27. September 1995, Braunschweig

Meeting on Statistics in Ecotoxicology
26-27 September 1995, Braunschweig

Bearbeitet von
Compiled by

Gerd Joermann
Herbert Köpp
Christine Kula

Fachgruppe Biologische Mittelprüfung
Abteilung für Pflanzenschutzmittel und Anwendungstechnik
Koordinierungsgruppe

Herausgeber

Biologische Bundesanstalt für Land- und Forstwirtschaft
Braunschweig, Deutschland



BBA

Verlag:

Eigenverlag

Vertrieb:

Saphir-Verlag, Gutsstraße 15, D-38551 Ribbesbüttel

Telefon 0 53 74/65 76

Telefax 0 53 74/65 77

ISSN-Nummer: 0947-8809

Kontaktadresse:

Dr. Gerd Joermann

Biologische Bundesanstalt für Land- und Forstwirtschaft

Fachgruppe für Biologische Mittelprüfung

Messeweg 11/12

D-38104 Braunschweig

Telefon +49/(0)5 31/2 99 3613

Telefax +49/(0)5 31/2 99 30 05

© Biologische Bundesanstalt für Land- und Forstwirtschaft

Das Werk ist urheberrechtlich geschützt. Die dadurch begründeten Rechte, insbesondere die der Übersendung, des Nachdrucks, des Vortrages, der Entnahme von Abbildungen, der Funksendung, der Wiedergabe auf fotomechanischem oder ähnlichem Wege und der Speicherung in Datenverarbeitungsanlagen, bleiben, auch bei nur auszugsweiser Verwertung, vorbehalten.

Biologische Bundesanstalt für Land- und Forstwirtschaft

Fachgespräch zur Statistik in der Ökotoxikologie

26. - 27. September 1995, Braunschweig

Gliederung

1. Einführung
2. Überblick über Aktivitäten in der OECD (H. Köpp)
3. Übersicht über das Versuchsdesign ökotoxikologischer Testmethoden (C. Kula)
4. Diskussion zu allgemeinen Themen
5. Ist die No-Observed-Effect Concentration (NOEC) eine geeignete statistische Kenngröße im ökotoxikologischen Test? (H. T. Ratte)
6. ECx-Schätzung als Alternative zum NOEC-Konzept
 - 6.1 Nachteile des NOEC-Konzepts
 - 6.2 Grundlagen der ECx-Schätzung
 - 6.3 Beispiele von ECx-Auswertungen (W. Pestemer, C. Kula, M. Streloke, G. Joermann)
 - 6.4 Diskussion
7. Schlußfolgerungen und Empfehlungen
8. Bibliographie

Anhang

Federal Biological Research Centre for Agriculture and Forestry

Meeting on Statistics in Ecotoxicology

26-27 September 1995, Braunschweig

Contents

1. Introduction
2. Summary of OECD activities (H. Köpp)
3. The design of ecotoxicological tests (C. Kula)
4. Discussion of general topics
5. Is the No-Observed-Effect Concentration (NOEC) a suitable statistical parameter in ecotoxicological tests? (H.T. Ratte)
6. ECx estimate as an alternative to the NOEC approach
 - 6.1 Disadvantages of the NOEC approach
 - 6.2 Basis ideas of ECx estimate
 - 6.3 Examples of ECx evaluations (W. Pestemer, C. Kula, M. Streløke, G. Joermann)
 - 6.4 Discussion
7. Conclusions and recommendations
8. Bibliography

Appendix

1. Einführung

Die statistische Auswertung ökotoxikologischer Versuche gewinnt zunehmend an Bedeutung. Ergebnisse können nur dann für die Bewertung von Pflanzenschutzmitteln und Chemikalien und für Entscheidungen herangezogen werden, wenn ihre Aussagekraft statistisch untermauert ist.

Bei vielen derzeit gebräuchlichen Testmethoden fehlen in den Richtlinien Hinweise zur statistischen Analyse der Daten. Dieser Mangel wird besonders problematisch, wenn die Rohdaten in einer Struktur erzeugt werden, auf die sich statistische Methoden nur schwer anwenden lassen. Deshalb muß schon bei der Entwicklung des Versuchsdesigns die Auswertung berücksichtigt werden.

Die Abteilung für Pflanzenschutzmittel und Anwendungstechnik der Biologischen Bundesanstalt ist an der Entwicklung internationaler Richtlinien in der Ökotoxikologie maßgeblich beteiligt. Um Fragen des Versuchsdesigns und der Auswertung sachkundig vertreten zu können, wurde am 26. und 27. September 1995 ein Fachgespräch durchgeführt, zu dem neben Kollegen aus Instituten der BBA auch externe Fachleute hinzugezogen wurden.

Ein wichtiges Ziel war dabei, die verschiedenen Prüfbereiche - Aquatik, Nutzorganismen, Bodenfauna, Vögel, etc. - an einen Tisch zu bringen. Gerade in der terrestrischen Ökotoxikologie wird häufig beklagt, daß die Tests und die Prüfstrategien so heterogen sind. Es wäre unglücklich, wenn diese Kluft nun durch unterschiedliche statistische Konzepte noch größer würde. Deshalb ist der Erfahrungsaustausch zwischen den Prüfbereichen besonders wichtig.

Auch auf internationaler Ebene ist die Statistik in der Ökotoxikologie in der Diskussion. Die OECD hat 1993 die Studie "A review of statistical data analysis and experimental design in OECD aquotoxicity test guidelines" (Pack, 1993) erstellen lassen. Darin werden nicht nur einzelne Testrichtlinien kritisch betrachtet, sondern es werden auch statistische Routineverfahren, wie sie derzeit gebräuchlich sind, in Frage gestellt. Ein Vorschlag ist, ANOVA-Methoden zur Bestimmung von NOEC-Werten durch Modellierung von Dosis-Wirkungs-Kurven und EC-Punkt-Schätzungen zu ersetzen. Dies berührt nicht nur die Auswertung der Versuche, sondern hätte weitreichende Auswirkungen einerseits auf das Versuchsdesign und andererseits auf die Umsetzung der Ergebnisse in Bewertungen. Die OECD veranstaltet im zweiten Halbjahr 1996 einen Workshop zu diesem Thema, der von der BBA in Braunschweig ausgerichtet wird. Das Fachgespräch war deshalb auch als eine Vorbereitung für diese OECD-Veranstaltung gedacht.

Damit sollte dieses Fachgespräch im wesentlichen dazu dienen

- die vorhandenen Informationen über diese wichtige Entwicklung zu streuen
- den derzeitigen Stand der statistischen Auswertung in allen ökotoxikologischen Prüfbereichen sowie mögliche Konsequenzen des Gutachtens zu diskutieren
- soweit wie möglich einen Konsens für weitere nationale Stellungnahmen und für die OECD-Tagungen vorzubereiten.

Im Rahmen des Fachgesprächs gab es drei Vorträge zu abgeschlossenen Themen, die in diesem Heft in eigenen Kapiteln erscheinen. Ansonsten wurden die Themen jeweils durch knappe Vorträge eingeführt und dann ausführlich diskutiert. Es war deshalb sinnvoll, auf die Wieder-

gabe der einzelnen Kurzreferate zu verzichten und stattdessen die Ergebnisse der Diskussion im Zusammenhang darzustellen.

Die Organisatoren möchten sich an dieser Stelle bei allen Teilnehmern recht herzlich bedanken. Ihre Fachkenntnis und Diskussionsfreude haben die Veranstaltung zum Erfolg werden lassen.

Gerhard Joermann
Herbert Köpp
Christine Kula

2. Überblick über Aktivitäten in der OECD (H. Köpp)

Pflanzenschutzmittel werden seit langem mittels standardisierter Testverfahren auf die im Zulassungsverfahren zu beurteilenden Aspekte (z.B. Wirksamkeit, Verbleib in der Umwelt, Rückstandsverhalten, Toxikologie und Ökotoxikologie) hin untersucht. Viele dieser Richtlinien sind bereits international harmonisiert, um die weitgehend identischen Datenanforderungen vieler Zulassungsbehörden mit möglichst wenig verschiedenen Studien erfüllen zu können. Dies trägt zur Minimierung des Einsatzes von Versuchstieren, Labor- und Finanzressourcen bei.

Die in der Toxikologie und Ökotoxikologie seit längerem eingesetzten Testrichtlinien sind überwiegend durch die OECD (Organisation for Economic Cooperation and Development) harmonisiert worden. Für Richtlinien, die erst in den letzten Jahren national oder in anderen Organisationen (ISO = International Standardization Organization, IOBC = International Organisation for Biological Control) entstanden sind, wird dies in den nächsten Jahren erfolgen.

Das OECD-Testrichtlinienprogramm ist das offizielle Verfahren zur Harmonisierung von Testrichtlinien in der OECD. Es wird finanziell auch von der EG-Kommission unterstützt. Anregungen für neue Richtlinien oder die Überarbeitung bestehender Methoden kommen von Mitgliedsstaaten, internationalen Organisationen (WHO, EU; ECPA, SETAC); oder aus OECD-Gremien („Pesticide Forum“ für Pflanzenschutzmittel; „Joint Meeting“ für Industriechemikalien; OECD-Sekretariat). Das Programm wird durchgeführt von den „National Coordinators“ aller Mitgliedsländer (für Deutschland: Dr. A.W. Lange, Umweltbundesamt) in Zusammenarbeit mit dem Sekretariat. Die „National Coordinators“ diskutieren Anregungen und Entwürfe, erstellen Prioritäten und ein Arbeitsprogramm, vergeben Arbeitsaufträge, verteilen Entwürfe zur Kommentierung an nationale Experten, erstellen daraus eine nationale Stellungnahme sowie die Endfassung eines Entwurfs. Arbeitsprogramm und Endfassungen der Richtlinienentwürfe bedürfen formell der Zustimmung des „Joint Meeting“ in gemeinsamer Sitzung mit dem „Pesticide Forum“.

OECD-Mitgliedsländer sind durch das „Mutual Agreement on Acceptance of Data“ (MAD) verpflichtet, nach OECD-Methoden durchgeführte Tests anzuerkennen, soweit sie ordnungsgemäß, das heißt ohne ungerechtfertigte Abweichung von Methodik oder Qualitätskriterien, durchgeführt wurden. Prinzipien zur GLP sind ebenfalls in der OECD verabschiedet.

Ausgelöst wurde das OECD-Testrichtlinienprogramm durch den Bedarf für Standardmethoden zur Beurteilung von Industriechemikalien Ende der 60er Jahre. An den Datenanforderungen und Beurteilungskriterien für diese Stoffe (siehe Chemikaliengesetz für Deutschland) orientierte sich auch das Testrichtlinienprogramm der OECD und damit die vorhandenen Methoden und ihr Versuchsdesign. Die Ökotoxikologie im Zulassungsverfahren für Pflanzenschutzmittel hat sich jedoch seitdem wesentlich schneller und weiter entwickelt, so daß aus heutiger Sicht dieser geschichtliche Hintergrund zu folgenden Mängeln führte, die die existierenden OECD-Testrichtlinien aufweisen:

- Mehrere OECD-Richtlinien sind in Design und statistischer Auswertung auf die Ermittlung von LC50/LD50-Werten ausgerichtet, da diese Kenngrößen für die Einstufung und Kennzeichnung von Industriechemikalien erforderlich sind. Für die Zulassungskriterien von Pflanzenschutzmitteln sind jedoch NOEC/NOEL-Werte inzwischen wesentlich wichtiger.
- Richtlinien zur Ermittlung von NOEC/NOEL benutzen zur statistischen Absicherung dieser Werte Signifikanztests, in die nur die Werte der Kontrolle sowie maximal zwei der Dosis-

stufen eingehen. Werte im Wirkungsbereich werden nicht weiter statistisch verarbeitet, obwohl in der Bewertung zunehmend auch eine Quantifizierung der Effekte bei Überschreitung der NOEC/NOEL erforderlich wird (Beurteilung der Effizienz von Risikominderungsmaßnahmen; Wichtung des Ausmaßes der Effekte in einer Nutzen-Risiko-Abwägung).

- Signifikanztests zur Absicherung der NOEC/NOEL leiden unter hoher Varianz der Daten, insbesondere der Kontrolldaten. Hohe Varianz führt häufig dazu, daß eine statistische Absicherung einer NOEC/NOEL erst bei hohen Dosisstufen mit bereits starken Effekten in der LOEC/LOEL möglich ist. Da hohe Variabilität unter anderem durch unsaubere Versuchsdurchführung und geringe Anzahl von Wiederholungen verursacht wird, belohnt diese Auswertungsweise eine suboptimale Versuchsdurchführung des Antragstellers letztlich durch höhere NOEC-Werte und damit leichteres Passieren der Zulassungskriterien (die in der Regel als Quotient aus der erwarteten Exposition und der NOEC ausgedrückt werden). Dieses Problem wird noch dadurch verschärft, daß insbesondere ältere Testrichtlinien keine Qualitätsmerkmale für die Kontrolle enthalten.
- Bei der Festlegung einer NOEC/NOEL mittels Signifikanztests ist die NOEC zwangsläufig identisch mit einer der Dosisstufen. Die NOEC ist deshalb durch den Versuchsansteller beeinflussbar, z.B. durch Wiederholung des Tests mit geänderten Dosisstufen. Auch die Abstufung zwischen den einzelnen Dosisstufen spielt eine Rolle.

Diese Nachteile sind vor dem Hintergrund zu sehen, daß die Prüfung von Pflanzenschutzmitteln die Behörden vieler Länder vor große Kapazitätsprobleme stellt. Lange Bearbeitungszeiten vor einer Zulassung wiederum benachteiligen Industrie und Landwirtschaft und können zu Wettbewerbsverzerrungen führen. Deshalb wird angestrebt, die Wirkstoffberichte der Zulassungsbehörden so transparent zu gestalten, daß sie gegenseitig anerkannt und als Basis der eigenen Entscheidung benutzt werden können. Die EU hat diese gegenseitige Anerkennung von Auswertungen bereits in den Regelungen zur Aufnahme von Wirkstoffen in Anhang I der Richtlinie 91/414/EWG vorgesehen. An den Vorgaben für die Ausgestaltung solcher Wirkstoffberichte (Monographien) wird seit Mitte 1994 intensiv gearbeitet. Im Rahmen der OECD wird ab Anfang 1996 versucht werden, auf Basis des EU-Entwurfes einen OECD-weiten Standard zu entwickeln. Dieser würde alle westlich orientierten Industrieländer und damit ca. 2/3 des Weltmarktes an Pflanzenschutzmitteln umfassen.

Die bisherigen Erfahrungen haben gezeigt, daß die Anerkennung der Auswertungen anderer Behörden nur dann gelingt, wenn der Berichtersteller neben vielen Details auch angibt, wie eine bestimmte Schlußfolgerung, z.B. die Festlegung und statistische Absicherung einer NOEC, entstanden ist. Durch eine solche internationale Arbeitsteilung kann die Bearbeitungszeit erheblich verkürzt, sowie eine höhere Konsistenz der Entscheidungen in verschiedenen Ländern erreicht werden. Sie hat jedoch für den einzelnen Bewerter den Nachteil, daß er für seine Entscheidung auf die Angaben in den Berichten einer anderen Behörde angewiesen ist, anstatt wie bisher unmittelbar Zugang zu den umfangreichen Rohdaten zu haben. Damit ist es in der Regel nicht mehr möglich, gegebenenfalls eigene statistische Auswertungen durchzuführen. Dieser Umstand verstärkt die oben geschilderten Nachteile der bestehenden Testrichtlinien deutlich.

Die OECD hat aus diesen Gründen ein Gutachten zur statistischen Auswertung von Tests zur aquatischen Ökotoxikologie erstellen lassen. Dieses Gutachten (Pack, 1993) bestätigt die Kritikpunkte und empfiehlt, anstelle der Festlegung einer NOEC/NOEL mittels Signifikanztests eine Modellierung der Dosis-Wirkungs-Kurve zur Ermittlung eines Punktes nahe der Wirkungsschwelle (z.B. EC05). Zum optimalen Einsatz dieser EC-Punkt-Schätzung wäre eine Änderung des Versuchsdesigns erforderlich (weniger Wiederholungen, dafür mehr Dosisstu-

fen, insgesamt möglicherweise weniger Versuchstiere), was weitreichende Folgen auf das Testrichtlinienprogramm hätte. Das Gutachten ist seit der Publikation intensiv diskutiert worden. In den Niederlanden und in Großbritannien wurden Tagungen dazu durchgeführt (The Hague, 1994; SETAC UK, April 1995). Die OECD-Mitgliedsländer haben über ihre nationalen Koordinatoren Kommentare eingereicht. Die Resonanz war überwiegend positiv, so daß die OECD diese Vorschläge weiter verfolgen wird. Dazu sind folgende Schritte vorgesehen:

1. Auswertung vorhandener Ringtestergebnisse (Daphnien-Reproduktionstest; Fisch-28 Tage Wachstumstest) mit Signifikanztests und ECx-Verfahren als Diskussionsgrundlage
2. Durchführung einer OECD-Tagung auf wissenschaftlich-technischem Niveau (Biostatistiker, Ökotoxikologen, Bewerter) zum Pack-Gutachten
3. Wenn die Expertentagung die Umsetzung des Gutachtens empfiehlt, soll ein weiteres Treffen auf höherer Ebene (Bewerter und Entscheidungsträger in Zulassungsverfahren) die Konsequenzen und eine Umsetzungsstrategie beraten.

Da das Gutachten auf aquatische Tests beschränkt wurde, muß die Übertragbarkeit auf die anderen Prüfbereiche (Vögel, Boden- und Nutzorganismen) diskutiert werden. Damit sind auch andere Organisationen betroffen, die Richtlinien erarbeiten (ISO, IOBC). Bezüglich der Vorgaben für die Erstellung von Auswertungsberichten muß geprüft werden, ob die Ergebnisse aller Dosisstufen aufgeführt werden müssen. Ferner ist zu beraten, ob (und wie) bei Ablösung der NOEC durch z.B. eine EC05 die in bestehenden Bewertungskonzepten (z.B. Uniform Principles der EU) benutzten Trigger und Sicherheitsfaktoren zu ändern sind.

Zur Umsetzung im Testrichtlinienprogramm bestehen zwei Optionen:

- jede einzelne Testrichtlinie mit der statistischen Auswertungsmethode zu ergänzen
- eine separate "Auswertungsrichtlinie" zu erstellen und die Empfehlungen zur Statistik aus den einzelnen Richtlinien zu streichen (diese Option wird von vielen bevorzugt, da sie schneller und flexibler ist).

3. Übersicht über das Versuchsdesign ökotoxikologischer Testmethoden (C. Kula)

Im folgenden werden international standardisierte Tests betrachtet, die bei der Prüfung von Pflanzenschutzmitteln und Chemikalien routinemäßig durchgeführt werden. Diese Tests lassen sich in Versuche mit qualitativen (quantalen) Meßgrößen und solche mit quantitativen (metrischen) Meßgrößen unterteilen. Der Einfachheit halber wird bei der Beschreibung der Expositionshöhe nur von "Dosis" gesprochen, obwohl es sich in vielen Versuchen um Konzentrationen handelt.

Versuche mit qualitativen Meßgrößen

Tabelle 1 gibt einen Überblick. Die Meßgröße ist in der Regel die Mortalität. In vielen dieser zumeist durch die OECD standardisierten Versuche werden 5 Dosisstufen eingesetzt. Die Abstufung der Dosen ist in den Richtlinien nicht festgelegt. Üblich sind Faktoren zwischen 1.78 und 3.16, entsprechend zwei bis vier Stufen pro Zehnerpotenz. Der Versuchsumfang liegt insgesamt zwischen 10 und 40 Individuen pro Dosisstufe. Bei Versuchen mit Vögeln und Fischen werden alle Tiere einer Dosisstufe in einen Käfig bzw. ein Aquarium eingesetzt, so daß die Frage der Unabhängigkeit der einzelnen Individuen besonders zu beachten ist. Bei anderen Versuchen werden die Tiere einer Dosisstufe auf mehrere Gefäße aufgeteilt, wobei dies bei der Auswertung aber in der Regel nicht berücksichtigt wird.

Tabelle 1: Versuche mit qualitativen Meßgrößen im Zulassungsverfahren

Test	Anzahl Dosisstufen	Abstufungsfaktor der Dosisreihe	Umfang (Individuen/Dosisstufe)	Zielgröße	Statistische Methode
<i>Vögel akut oral</i>	5	üblich 1.78	1 x 10	LD ₅₀	Probit
<i>Vögel subakut (5 d)</i>	5	üblich 1.78	1 x 10	LC ₅₀	Probit
<i>Fisch akut (96 h)</i>	5	üblich 3.16	1 x 10	EC ₅₀	Probit
<i>Daphnie akut (48 h)</i>	5	üblich 3.16	4 x 5	EC ₅₀	Probit
<i>Biene akut oral (48 h)</i>	nicht festgelegt	nicht festgelegt	3 x 10	LD ₅₀	Probit
<i>Biene akut Kontakt (48 h)</i>	nicht festgelegt	nicht festgelegt	3 x 10	LD ₅₀	Probit
<i>Regenwurm akut (14 d)</i>	5	üblich 1.78	4 x 10	LC ₅₀	Probit

Wichtigstes Versuchsziel ist einheitlich die Bestimmung der LD₅₀/LC₅₀, wobei die Mortalitätsrate jeweils mit Hilfe der Probit-Analyse in Abhängigkeit von der logarithmierten Dosis modelliert wird. Die Art der Auswertung wird nicht in Frage gestellt, wohl aber kann man darüber diskutieren, ob die LD₅₀/LC₅₀ der geeignete Parameter für Risikoabschätzungen ist.

In einigen Tests, z.B. mit Vögeln und Regenwürmern, werden neben der Mortalität auch quantitative Merkmale wie Körpergewicht und Futterkonsum erfaßt. Teilweise lassen sich auf

Basis dieser quantitativen Merkmale NOEC-Werte ableiten. Allerdings ist dies in den Richtlinien nicht ausdrücklich vorgesehen, und die Wahl der Dosisstufen erfolgt allein im Hinblick auf die Ermittlung der LD50/LC50, was dazu führt, dass die NOEC für Körpergewicht und Futterkonsum häufig unterhalb der niedrigsten getesteten Dosis liegt.

Versuche mit quantitativen Meßgrößen

Bei Versuchen mit quantitativen Meßgrößen handelt es sich meist um längerfristige Versuche (Tabelle 2). Merkmale sind z.B. Körpergewicht oder Reproduktionsrate. Generell gibt es bei diesen Versuchen hinsichtlich des Versuchsdesigns zwei Konzepte. Das eine sieht ausdrücklich die Bestimmung einer NOEC vor. Dazu werden 3 bis 5 Dosisstufen so gewählt, daß zumindest in der untersten Stufe kein statistischer Unterschied zur Kontrolle auftritt. Die Wahl der Dosisstufen wird aufgrund von Erfahrungswerten und Vorversuchen vorgenommen. Dieses Konzept wird bei aquatischen Organismen und Vögeln verfolgt. Bei Versuchen mit Nutzorganismen, der Bodenmikroflora und im Regenwurm-Reproduktionstest wird dagegen die Stärke des Effekts bei praxisrelevanter Exposition ermittelt. Hier werden eine bzw. zwei Dosisstufen gewählt, die dem üblichen Aufwand von Pflanzenschutzmitteln (und ggf. einem Vielfachen davon) entsprechen. Nach dem ersten Konzept wird zu einer Effektstärke (NOEC) die zugehörige Dosis ermittelt, im zweiten Fall zu einer Dosis die zugehörige Effektstärke.

Tabelle 2: Versuche mit quantitativen Meßgrößen im Zulassungsverfahren

Test	Anzahl Dosisstufen	Abstufungsfaktor der Dosisreihe	Umfang (Individuen/ Dosisstufe)	Zielgröße	Statistische Methode
<i>Fisch längerfristig (21 d)</i>	5	üblich 3.16	1 x 10 Ind.	NOEC	ANOVA
<i>Daphnia chronisch (21 d)</i>	5	üblich 3.16	10 x 1/ 4 x 10 Ind.	NOEC	ANOVA
<i>Alge chronisch (72 h)</i>	5	üblich 3.16	3-6 Gefäße (6 Kontr.)	NOEC	ANOVA
<i>Vögel Reproduktion (ca. 20 Wo)</i>	3	2 - 10	12 Paare	NOEC	ANOVA
<i>Regenwurm Reproduktion (8 Wo)</i>	2	5	4 x 10 Ind.	% Effekt	ANOVA
<i>Bodenmikroflora (28-100 d)</i>	2	5	3 Gefäße	% Effekt	ANOVA
<i>Nutzorganismen</i>	1	-	je nach Art verschieden	% Effekt	ANOVA

Der Versuchsumfang pro Dosisstufe ist sehr unterschiedlich, je nach Organismen und den im Versuch erhobenen Merkmalen. Der Versuchsumfang hat jedoch entscheidenden Einfluß auf die Aussagekraft der Versuche (siehe Kapitel 5). In der Vergangenheit wurde dieser Zusammenhang bei der Entstehung der Richtlinien häufig nicht berücksichtigt. Vielmehr spielten praktische Erwägungen oder Zufälligkeiten eine Rolle. Bei der Entwicklung neuer Richtlinien sollte man hier, unter Hinzuziehung von Statistikern, planmäßiger vorgehen.

Die Richtlinien der Versuche mit mehreren Dosisstufen sehen die Bestimmung der NOEC vor, das heißt, der höchsten Testkonzentration, bei der kein signifikanter Unterschied zur Kontrolle vorliegt. Die statistische Auswertung erfolgt einheitlich über multiple Mittelwertvergleiche, z.B. Dunnett-Test, Williams-Test oder LSD-Test (im folgenden als ANOVA-Verfahren zusammengefaßt), wobei die Angaben in den Richtlinien unterschiedlich detailliert sind. Bis vor kurzem wurde diese Auswertung nicht in Frage gestellt. Daß bei qualitativen Merkmalen wie selbstverständlich Dosis-Wirkungs-Kurven modelliert werden (s.o.), bei quantitativen Merkmalen dagegen genauso selbstverständlich Mittelwertvergleiche zwischen Behandlung und Kontrolle durchgeführt werden, ist auffällig und läßt sich kaum vernünftig erklären.

4. Diskussion zu allgemeinen Themen

Versuchskonzepte mit Dosis-Reihen und Ein-Dosis-Tests

Die unterschiedliche Konzeption (Dosis-Reihen und Ein-Dosis-Tests) ist historisch bedingt (siehe Kapitel 3) und hat kaum einen fachlichen Hintergrund. Ein-Dosis-Tests wurden im Zusammenhang mit der Prüfung von Pflanzenschutzmitteln entwickelt, wo die praxisrelevante Exposition der auf der Kulturfläche lebenden Organismen besser absehbar ist als bei allgemeinen Chemikalien. Ein gravierender Nachteil ist in diesem Fall, daß die Ergebnisse eines solchen Ein-Dosis-Tests nur für ein bestimmtes Expositionsszenario verwendbar sind. Soll das Pflanzenschutzmittel mit einem höheren oder niedrigeren Aufwand angewendet werden, ist häufig eine Wiederholung des Tests nötig. Deshalb halten die Teilnehmer des Fachgesprächs die Testung von Dosis-Reihen, zumindest bei Standard-Labortests, für eindeutig geeigneter, und zwar unabhängig von der Frage, ob die Auswertung mit ANOVA oder ECx-Bestimmung erfolgt.

Validitätskriterien in den Testrichtlinien

Die Validität von Testergebnissen hängt entscheidend von der Qualität des Testsystems im jeweiligen Versuch ab (Gesundheitszustand der Versuchstiere, Haltungsbedingungen, etc.). Dies ist besonders bei ANOVA-Auswertungen zu beachten, da mit der Erhöhung der Varianz der β -Fehler, d. h. die Wahrscheinlichkeit, einen tatsächlichen vorhandenen Unterschied zwischen Kontrolle und Behandlung zu übersehen, zunimmt. Die Teilnehmer des Fachgesprächs halten es für erforderlich, daß in den Richtlinien geeignete Validitätskriterien definiert werden, z. B. Vorgaben zur Höchstgrenze der Mortalität in der Kontrolle, Mindestzahlen zur Reproduktionsrate, oder auch eine Obergrenze für die maximal zulässige Streuung eines Merkmals in der Kontrolle. In neueren Richtlinien ist das häufig schon realisiert, während für ältere noch ein Nachholbedarf besteht.

Behandlung von Ausreißern

Es spricht nichts dagegen, Ausreißer aus dem Datensatz zu streichen, wenn sich das Zustandekommen der Abweichung plausibel erklären läßt. Starke Abweichungen sollten Anlaß sein, zunächst das Testsystem zu prüfen, falls das im Nachhinein noch möglich ist. Eine automatische Ausreißerelimination per EDV-Programm ist abzulehnen.

5. Ist die No-Observed-Effect Concentration (NOEC) eine geeignete statistische Kenngröße im Biotest? (H.T. Ratte)

Die Grundproblematik der NOEC existiert aus zwei Gründen: Erstens soll sie als Basis zur Abschätzung einer NEC (No Effect Concentration, Safe-Level) im Freiland dienen (NOEC und NEC werden leider zuweilen als dasselbe angesehen). Zweitens hängt die Beobachtungsgenauigkeit für die NOEC von einer statistischen Hypothesen-Testung ab, die keine sicheren Aussagen erlaubt. Solche Hypothesentests sind ihrer Natur nach konservativ, d.h. sie favorisieren die sogenannte Nullhypothese (H_0 : \approx Es wird unterstellt, daß kein substanzspezifischer Effekt existiert). Die H_0 wird oft sehr lange beibehalten, auch wenn bereits ein deutlicher Unterschied zur Kontrolle besteht. Hierdurch entsteht das Risiko, tatsächlich existierende Effekte nicht zu erkennen. Eine Abhilfe dieser unbefriedigenden Situation könnte darin bestehen, den sogenannten β -Fehler (Fehler 2. Art) zu kontrollieren. Dies soll im folgenden an einem Beispiel kurz erläutert werden.

Hypothetisches Beispiel

In einem chronischen Daphnia-Test wurden auch Effekte auf die die Länge der Tiere untersucht (angegeben in Micrometer-Teilstrichen). Es wurden zwei verschiedene Testkonzentrationen und eine Kontrolle geprüft. Der Stichprobenumfang betrug jeweils 10 Tiere pro Testansatz. Folgende Stichprobenkennwerte wurden ermittelt:

Testansatz	Mittelwert Symbol	Länge (Teilstriche)	Standardab- weichung s
Kontrolle	\bar{x}_k	33	6
Konzentration 1	\bar{x}_{a1}	30	6
Konzentration 2	\bar{x}_{a2}	27	6

Bestimmung der NOEC

Um eine NOEC zu bestimmen, wird in der Regel (so auch hier) ein multipler t-Test (Dunnett- oder Williams-Test) durchgeführt. Um die NOEC-Problematik und die Rolle des β -Fehlers zu beleuchten, muß man wissen, wie ein t-Test funktioniert. Der Einfachheit halber werden diese Betrachtungen jedoch nicht für die multiplen, sondern für den allseits bekannten paarweisen t-Test durchgeführt. Sie sind jedoch prinzipiell auch auf multiple t-Tests übertragbar, doch ist hierfür kaum eine anschauliche Darstellung möglich.

t-Test, α - und β -Fehler

Die paarweisen und multiplen t-Tests bewerten, ob eine beobachtete Mittelwertsdifferenz zufällig aufgetreten oder aufgrund einer toxischen Wirkung zustande gekommen sein könnte.

Grundlage hierbei ist die t-Formel:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - \mu}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad \text{oder, da } \mu = 0: \quad t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad (1)$$

Die Formel 1 dient der Berechnung eines t-Wertes, der einer standardisierten Differenz entspricht und für den eine Wahrscheinlichkeit angegeben werden kann, eine so große oder eine noch größere Differenz zu erhalten (zugehörige Wahrscheinlichkeitsverteilung: t-Verteilung nach STUDENT).

Warum ist dies möglich? Aufgrund des Zentralen Grenzwertsatzes der Statistik kann abgeleitet werden, daß unendlich viele Differenzen je zweier Mittelwerte eine Wahrscheinlichkeitsverteilung bilden, die eine Normalverteilung mit dem Mittelwert $\mu = 0$ und der Standardabweichung $\sigma_{\bar{x}_1 - \bar{x}_2}$ ist (= Stichprobenverteilung der Differenzen zweier Mittelwerte). Die Standardabweichung berechnet sich nach Formel 2:

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \quad (2)$$

und wird über die in den zwei zu testenden Stichproben nach Formel 3 geschätzt.

$$\sigma_{\bar{x}_1 - \bar{x}_2} \approx s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \quad (3)$$

In der t-Formel (1) wird der Abstand zwischen beobachteter Mittelwertsdifferenz $\bar{x}_1 - \bar{x}_2$ und dem Mittelwert μ auf der Basis der Standardabweichung der Wahrscheinlichkeitsverteilung der Differenzen bewertet (d.h. durch diese dividiert; da $\mu = 0$, wird es in den t-Formeln meist nicht aufgeführt). Der so erhaltene t-Wert dient dazu, die Wahrscheinlichkeit für eine beobachtete Differenz (hier z.B. $\bar{x}_k - \bar{x}_{a2}$) oder eine noch größere zu berechnen (= Fläche rechts von diesem Wert in Abb. 1; siehe statistische Tabellenwerke oder Formeln, alternativ kann man den Stichproben-t-Wert auch mit dem für das gewählte Signifikanzniveau tabellierten t-Wert vergleichen). Die Wahrscheinlichkeit für den Stichproben-t-Wert und der kritische Vergleichs-t-Wert unterscheiden sich bei paarweisen und multiplen Tests, jedoch ist die Vorgehensweise beim Test identisch.

Unser Beispiel läßt zwei t-Vergleiche zu, wobei zwei Differenzen gebildet werden können

$$\bar{x}_k - \bar{x}_{a1} = 3 \quad (4)$$

$$\bar{x}_k - \bar{x}_{a2} = 6 \quad (5)$$

Drei Teilstriche Differenz sind 9% des Mittelwerts der Kontrolle (= 33 Teilstriche), sechs Teilstriche 18% (d.h. ein 9%iger oder 18%iger Effekt). Außerdem ergibt sich, da $n_k = n_{a1,a2} = 10$, als beobachtete Standardabweichung $s_{\bar{x}_k - \bar{x}_{a1,a2}} = 2.683$ für beide Vergleiche.

Abb. 1 zeigt die Wahrscheinlichkeitsverteilung der Differenzen, die genau für unser Beispiel berechnet wurde, wobei der Anschaulichkeit halber die X-Achse in Einheiten der Meßgröße (Teilstrichen) angegeben ist. Eine solche Verteilung kommt rein zufällig dadurch zustande, daß je zwei Stichproben aus derselben Grundgesamtheit zufällig gezogen und die Differenzen der beiden Mittelwerte berechnet werden (für unendlich viele Versuche). Wir erkennen, daß sehr viele dieser Mittelwertsdifferenzen klein sind und nahe bei 0 liegen (Abb. 1), es gibt aber auch Fälle, in denen Differenzen bis zu 10 Teilstrichen auftreten (relativ zu den 33 Teilstrichen der

Kontrolltiere sind dies 30%, 30%iger Effekt). Differenzen dieser Größe können wohlgerne nicht durch Substanzwirkung, sondern rein zufällig zustande kommen.

Ab wann ist eine Differenz statistisch signifikant? Die beiden Differenzen aus unserem Beispiel sind auf der Abszisse eingetragen. Für beide nehmen wir zunächst an, daß sie auch zufällig zustande gekommen sein könnten (= H_0). Die Wahrscheinlichkeit für die erste Differenz $\bar{x}_k - \bar{x}_{a1}$ oder eine größere liegt bei etwa 0.13, für die zweite $\bar{x}_k - \bar{x}_{a2}$ oder eine größere bei etwa 0.01. Wir haben jedoch keinen Beweis dafür, daß nur der Zufall im Spiel war. Es könnte auch eine Substanzwirkung vorliegen, was aber nicht zu beweisen ist.

Wie ist zu entscheiden? Man hat eine Konvention getroffen, sogenannte Signifikanzniveaus einzuführen ($\alpha = 0.05$; $\alpha = 0.01$; $\alpha = 0.001$), wovon das gebräuchlichste das 5%-Niveau darstellt ($\alpha = 0.05$).

Der größere schraffierte Bereich in Abb. 1 entspricht $\alpha = 0.05$. Die Signifikanzgrenze teilt die Wahrscheinlichkeitsverteilung in zwei Bereiche ein. Im Bereich links von der Signifikanzgrenze behaupten wir, es liege keine Substanzwirkung vor („alles Zufall“), im Bereich rechts davon, es liege ein Substanzeffekt vor. Es wird hier die einseitige Fragestellung betrachtet. Obwohl die Richtung einer Substanzwirkung auf die Länge nicht unbedingt eine Verkürzung sein muß, so entspricht die einseitige Fragestellung beim statistischen Test meistens auch der Fragerichtung im Biotest (liegt eine Hemmung, Reduktion, Verkleinerung einer Größe vor?).

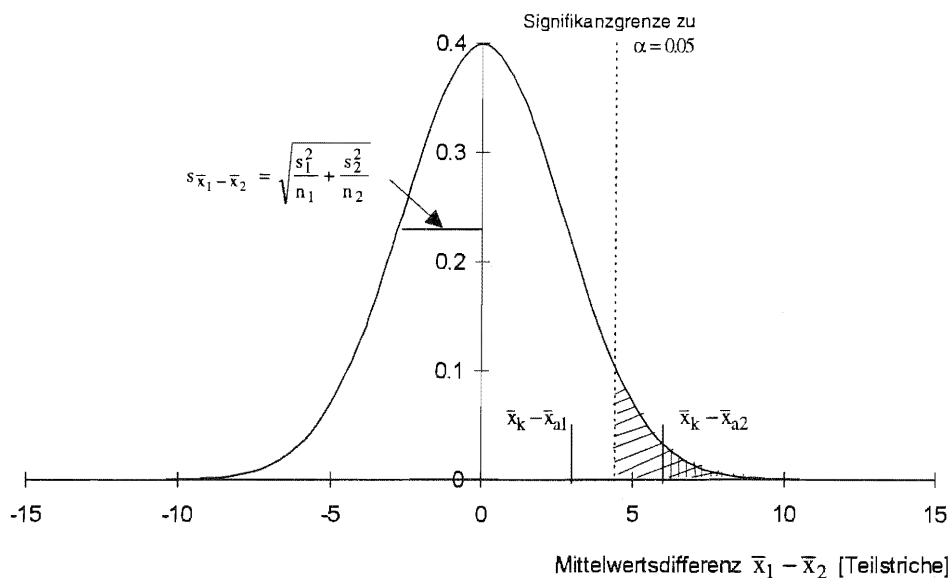


Abb. 1: Stichprobenverteilung der Differenzen zweier Mittelwerte für Übungsbeispiel n_k , n_{a1} und $n_{a2} = 10$. Die Verteilungskurve repräsentiert alle möglichen zufälligen Differenzen der Mittelwerte aus zwei Stichproben $\bar{x}_1 - \bar{x}_2$.

$\bar{x}_k - \bar{x}_{a1}$; $\bar{x}_k - \bar{x}_{a2}$: beobachtete Differenzen in unserem Beispiel (Erklärung siehe Text)

Nach dieser Datenlage würden wir die kleinere Differenz ($\bar{x}_k - \bar{x}_{a1}$) als mit der Nullhypothese vereinbar ansehen (= vermutlich kein Effekt), die größere Differenz ($\bar{x}_k - \bar{x}_{a2}$) wäre signifikant (= kommt rein zufällig selten vor). Hier unterstellt man einen Effekt und geht das Risiko ein, in weniger als 5% (bezogen auf das gewählte Niveau) oder genauer 1% (bezogen auf die tatsächliche Differenz) eine falsche Behauptung aufgestellt zu haben (Irrtumswahrscheinlichkeit, α -Fehler). Die NOEC wäre die Konzentration, deren Mittel-

wertsdifferenz gerade noch nicht signifikant ist (hier die Konzentration 1, die der Differenz $\bar{x}_k - \bar{x}_{a1}$ zugehört).

Hätte man dieselben Ergebnisse mit Stichprobengrößen von $n = 25$ erzielt, so wäre die Wahrscheinlichkeitsverteilung schmäler und als Lowest-Observed-Effect Concentration (LOEC) hätte die kleinere Differenz $\bar{x}_k - \bar{x}_{a1}$ statistisch abgesichert werden können ($p \leq 0.05$; Abb. 2). Eine NOEC wäre dann nicht zu bestimmen gewesen. Man erkennt daran, daß man die NOEC/LOEC durch Veränderung der Stichprobengrößen beliebig verschieben kann.

Wir kehren zurück zur Stichprobengröße n_k, n_{a1} und $n_{a2} = 10$ (Abb. 1) und der kleineren Differenz $\bar{x}_k - \bar{x}_{a1}$. Wir nehmen an, ein Stoff würde tatsächlich nur einen geringen Effekt ausüben und eine kleine Mittelwertsdifferenz von 3 Teilstrichen in der Länge (= 9%-Effekt) verursachen, was man normalerweise nicht weiß. Es ergibt sich dann die Situation in Abb. 3. Zufällige Differenzen zweier Mittelwerte aus Kontrolle und Konzentration 1 würden eine Verteilung bilden, die den Mittelwert $\mu_1 = 3$ hat. Dies ist die gestrichelte Kurve in Abb. 3.

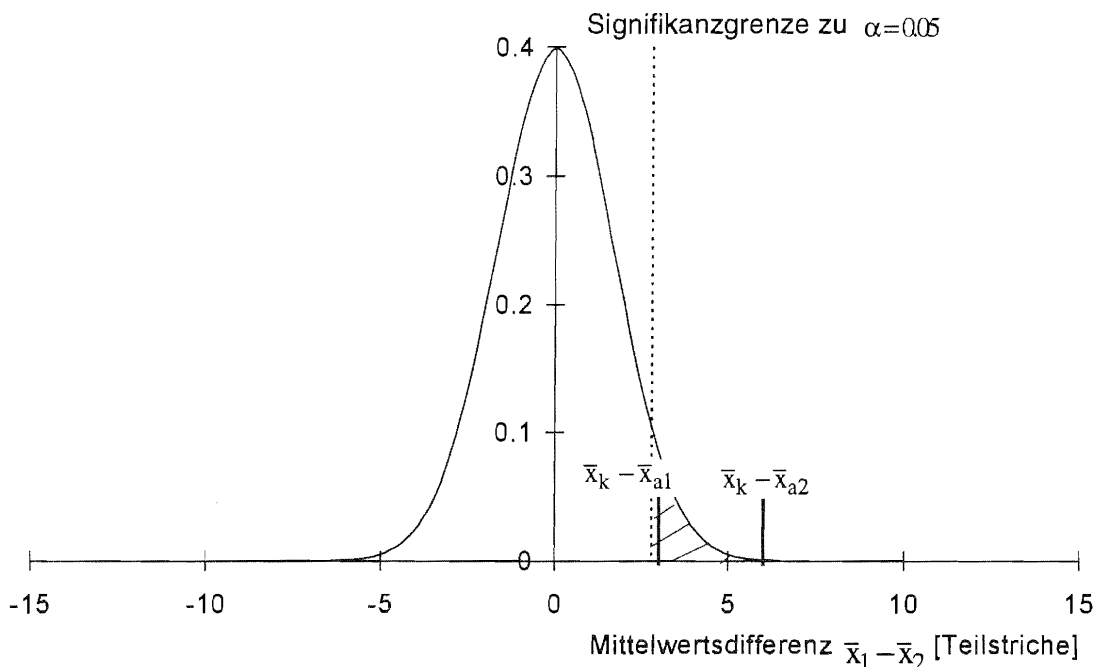


Abb. 2: Stichprobenverteilung der Differenzen zweier Mittelwerte für den Fall n_k, n_{a1} und $n_{a2} = 25$. Die Verteilungskurve repräsentiert alle möglichen zufälligen Differenzen der Mittelwerte aus zwei Stichproben $\bar{x}_1 - \bar{x}_2$.

$\bar{x}_k - \bar{x}_{a1}$; $\bar{x}_k - \bar{x}_{a2}$: beobachtete Differenzen in unserem Beispiel (Erklärung siehe Text)

Wir können nun zwei Wahrscheinlichkeiten angeben: Die Differenz $\bar{x}_k - \bar{x}_{a1}$ oder eine größere hat in der Verteilung mit $\mu = 0$ (kein Effekt) eine Wahrscheinlichkeit von ca. 0.13. Das ist das Flächenstück rechts von der in Abb. 3 eingezeichneten Differenz $\bar{x}_k - \bar{x}_{a1} = 3$. Gleichzeitig kann dieselbe Differenz $\bar{x}_k - \bar{x}_{a1}$ mit der zweiten gestrichelten Wahrscheinlichkeitsverteilung in Beziehung gebracht werden. Hier beträgt die Wahrscheinlichkeit dafür, daß die beobachtete Differenz $\bar{x}_k - \bar{x}_{a1}$ oder eine kleinere (!) erhalten wird, 0.50 (Flächenstück links von der in Abb. 3 eingezeichneten Differenz $\bar{x}_k - \bar{x}_{a1} = 3$ in der gestrichelten Verteilung). Diese Wahr-

scheinlichkeit wird als β -Fehler bezeichnet. Er wird normalerweise auf die Signifikanzgrenze bezogen, also wäre dann hier $\alpha = 0.05$ (Flächenstück rechts der Signifikanzgrenze, durchgezogene Verteilung mit $\mu = 0$) und β etwa 0.80 (Flächenstück links der Signifikanzgrenze, gestrichelte Verteilung mit $\mu = 3$). Bei dieser Konstellation würde man also die Nullhypothese (kein Effekt) beibehalten, obwohl die Wahrscheinlichkeit für einen Effekt deutlich größer ist.

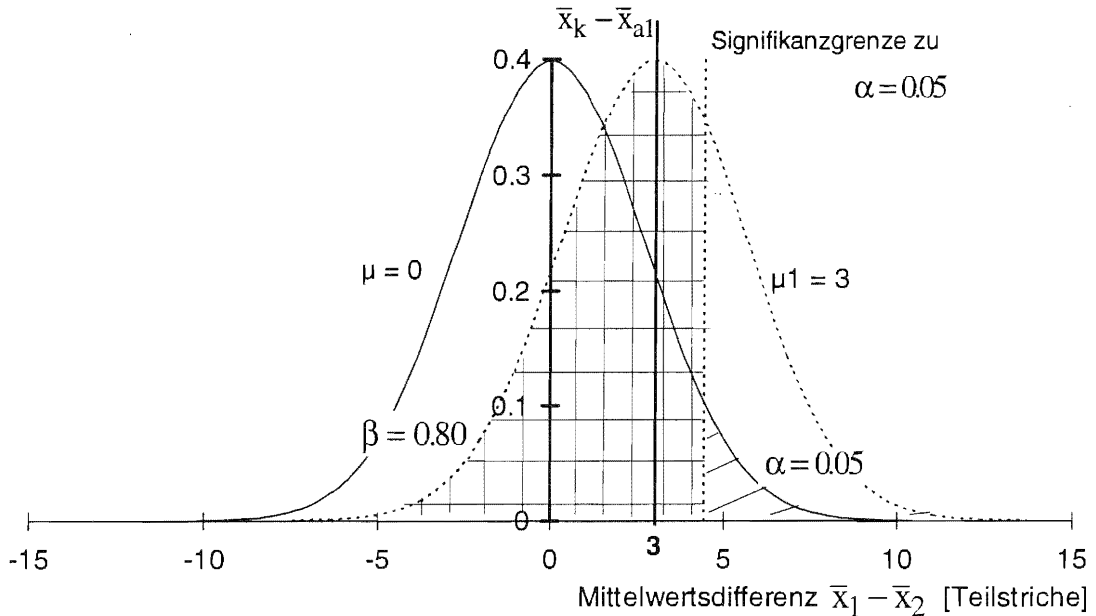


Abb. 3: Stichprobenverteilungen der Differenzen zweier Mittelwerte für Übungsbeispiel n_k und $n_{a1} = 10$. Die Verteilungskurve repräsentiert alle möglichen zufälligen Differenzen der Mittelwerte aus zwei Stichproben $\bar{x}_1 - \bar{x}_2$, wobei $\mu = 0$ (kein Effekt) und $\mu = 3$ (9%-Effekt, gestrichelt).

$\bar{x}_k - \bar{x}_{a1}$: beobachtete Differenz in unserem Beispiel (Erklärung siehe Text)

In der Praxis kennt man jedoch den „wahren Effekt“ nicht, weiß also nicht was die Substanz wirklich verursacht, und man ist allein auf Stichprobenergebnisse angewiesen. Man kennt somit weder die Größe von μ_1 noch den genauen β -Fehler. Wie kann man trotzdem den β -Fehler kontrollieren? Ideal wäre es, wenn der α - und β -Fehler etwa gleich gehalten werden könnte. Als tolerierbar gelten in Lehrbüchern der Statistik jedoch auch β -Fehler von 0.20. In der Sicherheitsforschung (z.B. Gentechnik) werden zuweilen „härtere“ Bedingungen gestellt: z.B. $\alpha = 0.10$ und $\beta = 0.05$.

Die Kontrolle des β -Fehlers durch den Experimentator ist nur durch die Wahl eines geeigneten Stichprobenumfangs möglich. Abb. 4 a zeigt, daß bei gleichem α - und β -Fehler eine Stichprobengröße von 10 in Kontrolle und Testansatz ausreichen würde, eine wahre Mittelwertsdifferenz von 9 Teilstrichen ($\mu_1 - \mu = 9$) oder mehr (27% Mindestunterschied zur Kontrolle) statistisch abzusichern ($p \leq 0.05$). Ließe man einen β -Fehler von 0.20 zu, so würde eine wahre Mittelwertsdifferenz von 7 Teilstrichen ($\mu_1 - \mu = 7$) oder mehr (21% Mindestunterschied zur Kontrolle) statistisch absicherbar sein ($p \leq 0.05$; Abb 4 b). Erhöhte man die Stichprobengröße auf 25, ließen sich Differenzen von 6 Teilstrichen (18% Mindestunterschied) statistisch absichern ($p \leq 0.05$; Abb. 4 c). Für die statistische Absicherung ($p \leq 0.05$) eines mindestens 11%igen Unterschieds zur Kontrolle ($\mu_1 - \mu = 3$ Teilstriche) wären jedoch 100 Replikate sowohl in Kontrolle als auch Testansatz notwendig (Abb. 4 c).

Die durchgeführten Betrachtungen und Rechnungen bezogen sich auf den paarweisen t-Test. Bei multiplen t-Tests würde die Absicherung einer bestimmten Differenz bei β -Fehler-Kontrolle noch mehr Replikate erfordern als beim paarweisen Test.

Welche Konsequenzen ergeben sich für die Praxis, welche Empfehlungen für den Bio-test?

Man sollte sich grundsätzlich folgende Vorgehensweise zu eigen machen, wenn man unbedingt die NOEC/LOEC berechnen und dies einer Bestimmung von Dosis-Wirkungs-Beziehungen (EC50, etc.) vorziehen will:

- Es ist festzulegen, welcher Effekt (= welche Mittelwertsdifferenz) biologisch bedeutsam sein dürfte.
- Danach bestimme man das Signifikanzniveau α für den Test (Irrtumswahrscheinlichkeit, d.h. die Wahrscheinlichkeit, daß man einen Effekt annimmt, obwohl keiner vorhanden ist).
- Wähle dann n so, daß der β -Fehler einen akzeptablen Wert annimmt (ideal: $\alpha = \beta$; max. tolerables $\beta = 0.20$).

Geht man so vor, so wird man erkennen, daß der Stichprobenumfang leicht die Grenzen des praktisch Machbaren übersteigt, selbst wenn der β -Fehler 0.80 beträgt. In der Praxis wird man somit selten Bedingungen antreffen, die eine solide NOEC-Bestimmung ermöglichen.

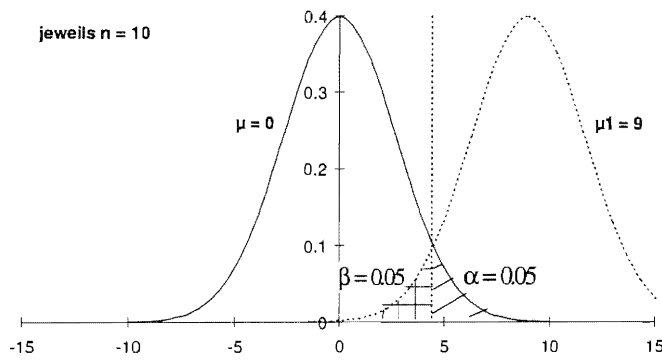
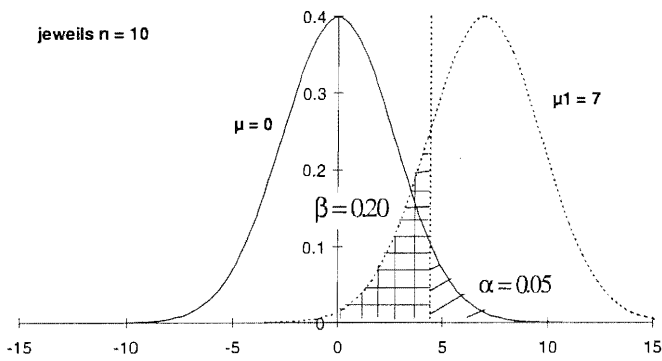
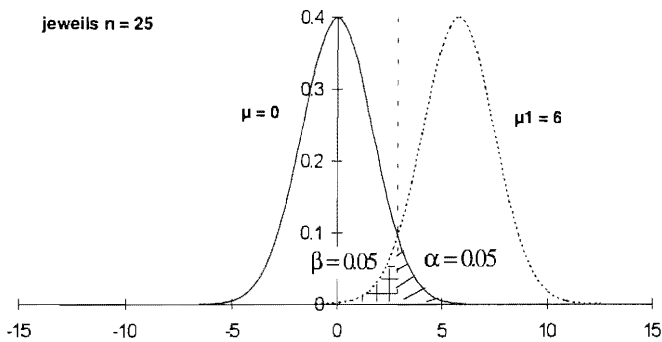
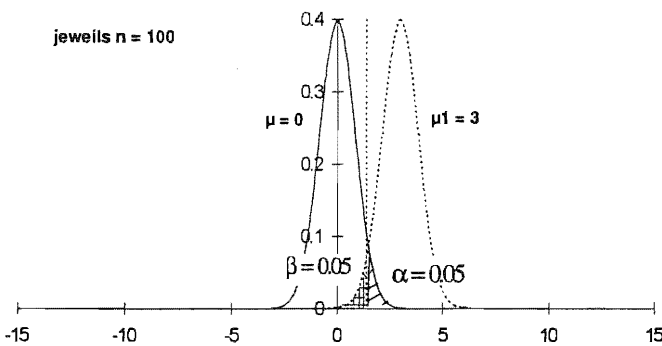
a**b****c****d**

Abb. 4: Verschiedene Konstellationen bei α -/ β -Fehlerkontrolle (Erklärung im Text)

Mittelwertsdifferenz $\bar{x}_1 - \bar{x}_2$ [Teilstriche]

6. ECx-Schätzung als Alternative zum NOEC-Konzept

6.1 Nachteile des NOEC-Konzepts

Die Bestimmung von NOEC-Werten mit Hilfe der ANOVA (und nachfolgenden Mittelwertvergleichen zwischen Behandlung und Kontrolle) hat eine Reihe von Nachteilen und Komplikationen. Die grundsätzlichen Probleme sind in dem Diskussionspapier herausgearbeitet, das Simon Pack für die OECD erstellt hat; sie wurden im Rahmen dieses Fachgesprächs von Dr. Ratte noch einmal erläutert (siehe Kapitel 5).

- Die NOEC ist zwangsläufig eine der Testkonzentrationen.
- Hohe Variabilität vermindert die Trennschärfe und führt damit zu höheren, d.h. „günstigeren“ NOEC-Werten.
- Es gibt keine Vertrauensbereiche für die NOEC-Werte
- Daten aus den Dosisstufen oberhalb der LOEC fließen nicht in das Ergebnis ein und werden praktisch verschwendet.

Grundsätzlich wurden diese Nachteile von allen Teilnehmern gesehen, wenn auch die Gewichtung unterschiedlich war. So wurde darauf hingewiesen, daß das Problem der hohen Variabilität teilweise durch Validitätskriterien gelöst werden kann. Außerdem sei der Bewerter einer Studie nicht unter allen Umständen an das Ergebnis der Signifikanztests gebunden. Eine deutliche Abweichung von der Kontrolle, die zwar nicht signifikant sei, aber unter Berücksichtigung der höheren Dosisstufen im Trend liege, könne durchaus als Effekt gewertet werden. Auf diese Weise würden dann doch auch die höheren Dosisstufen mitberücksichtigt.

6.2 Grundlagen der ECx-Schätzung

Eine Alternative zur ANOVA ist die Modellierung der Dosis-Wirkungs-Kurve und die Bestimmung einer bestimmten Effektkonzentration (EC-Punkt-Schätzung), also ein Verfahren, wie es auch bei qualitativen Merkmalen von Akuttests benutzt wird. Im Vergleich zu Akuttests ergeben sich allerdings zwei Probleme:

Erstens ist die Modellwahl für die quantitativen Merkmale langfristiger Tests nicht so unumstritten wie das Probitmodell für akute Mortalitätsdaten, und zweitens geht es bei langfristigen Tests um die Effektschwelle, d.h. um einen Bereich der Dosis-Wirkungs-Kurve, der grundsätzlich schwieriger zu modellieren ist als der Bereich der EC50.

Voraussetzung für eine solche Auswertung ist ein angepaßtes Testdesign. Im Gegensatz zur ANOVA-Auswertung müssen die Ressourcen auf mehr Dosisstufen mit jeweils kleinerem Umfang verteilt werden. Für die Anzahl der Replikate pro Dosisstufe gibt es keine Untergrenze, sofern der Gesamtumfang pro Versuch ausreichend groß ist. Die Anzahl der Dosisstufen sollte mindestens 4 betragen, besser jedoch 5 bis 6. Die Dosisstufen sollten möglichst in dem Bereich der Dosis-Wirkungs-Kurve liegen, der modelliert werden soll. Die Anforderungen an das range-finding sind deshalb sehr hoch. Bei einigen Teilnehmern besteht deshalb die Befürchtung, daß die gewählten Dosisstufen sich im Nachhinein als ungünstig herausstellen und keine Modellierung erlauben; eine NOEC lasse sich ersatzweise auch nicht bestimmen (wegen des

geringen Umfangs pro Dosisstufe), so daß der Versuch letztlich kein aussagefähiges Ergebnis liefere. Die Statistikfachleute hielten das Verfahren allerdings für so robust, daß auch bei einer nicht optimalen Datenlage durchaus noch eine Modellierung möglich sei. Ein weiterer Kritikpunkt bezog sich darauf, daß in der Regel nur ein Konzentrationsbereich eingestellt werden kann, der für eine Art und einen Endpunkt optimal ist. Sollen verschiedene Arten oder mehrere Endpunkte untersucht werden, ist die Auswahl der optimalen Testkonzentrationen schwierig.

6.3 Beispiele von ECx-Auswertungen

Pflanzen (W. Pestemer)

In Pflanzenwachstumstests hat sich die Modellierung von Dosis-Wirkungs-Kurven zur Standardauswertung entwickelt und wird als Anlage der vorläufigen BBA-Richtlinie (Teil VI, 13-1) zur „Prüfung der Phytotoxizität von Herbiziden auf nachgebaute Kulturen“ aufgeführt.

Prinzip der Testmethode

Für die Ermittlung der Dosis-Wirkungs-Beziehungen wird ein Wachstumstest verwendet. Dabei wird die Prüfsubstanz gleichmäßig in den Boden eingemischt und anschließend zur Einstellung des Sorptionsgleichgewichts über Nacht bei 4 °C gelagert. Es werden spezielle Dochtöpfe verwendet, die einen gleichmäßig hohen Wassergehalt des Bodens und damit eine sehr gute Pflanzenverfügbarkeit der Testsubstanz und hohe Empfindlichkeit der Testpflanzen gewährleisten. Am folgenden Tag wird der Boden in die Töpfe gefüllt und Keimlinge der Testpflanzen pikiert. Nach 10-14 Tagen Wachstum erfolgt die Bestimmung der Sproßfrischmasse. An diese Werte wird eine logistische Dosis-Wirkungs-Kurve angepaßt, der dann die geforderten Kenngrößen entnommen werden können. Für jeden Wirkstoff müssen zur Erstellung einer Dosis-Wirkungsbeziehungen mindestens 6 Konzentrationen plus unbehandelter Kontrolle in einem Test geprüft werden. Jede Konzentration muß in vier- bis sechsfacher Wiederholung angesetzt werden.

Auswertung

Eine genaue Sichtung der Daten vor der Verrechnung ist notwendig, da Tests mit biologischen Systemen meist eine große Variabilität zeigen. Bei sehr genauem Arbeiten und sorgfältiger Auswahl der Testpflanzen auf Einheitlichkeit kann die Variation innerhalb der Wiederholungen auf ca. 10 % gedrückt werden, doch ein Ausreißertest sollte trotzdem durchgeführt werden. Methoden sind zum Beispiel bei SACHS (1969) beschrieben. Weitere wichtige Größen, die zur Sichtung der Daten berechnet werden sollten, sind Mittelwert, Standardabweichung, Variationskoeffizient und der Anteil der Meßgröße prozentual zur unbehandelten Kontrolle.

Die beiden wichtigen Kenngrößen ED_{10} und ED_{50} werden aus der Dosis-Wirkungs-Beziehung bestimmt. Sie sind in der allgemeinen Darstellung in Abbildung 5 gezeigt. Die Frischmasse (FM) der Pflanzen ist hier gegen die logarithmierte Konzentration aufgetragen. ED_{10} und ED_{50} mit ihren Vertrauensbereichen können mit dem Programm LOGFIT oder anderen Möglichkeiten zur Anpassung logistischer Kurven (z.B. unter SAS) berechnet werden.

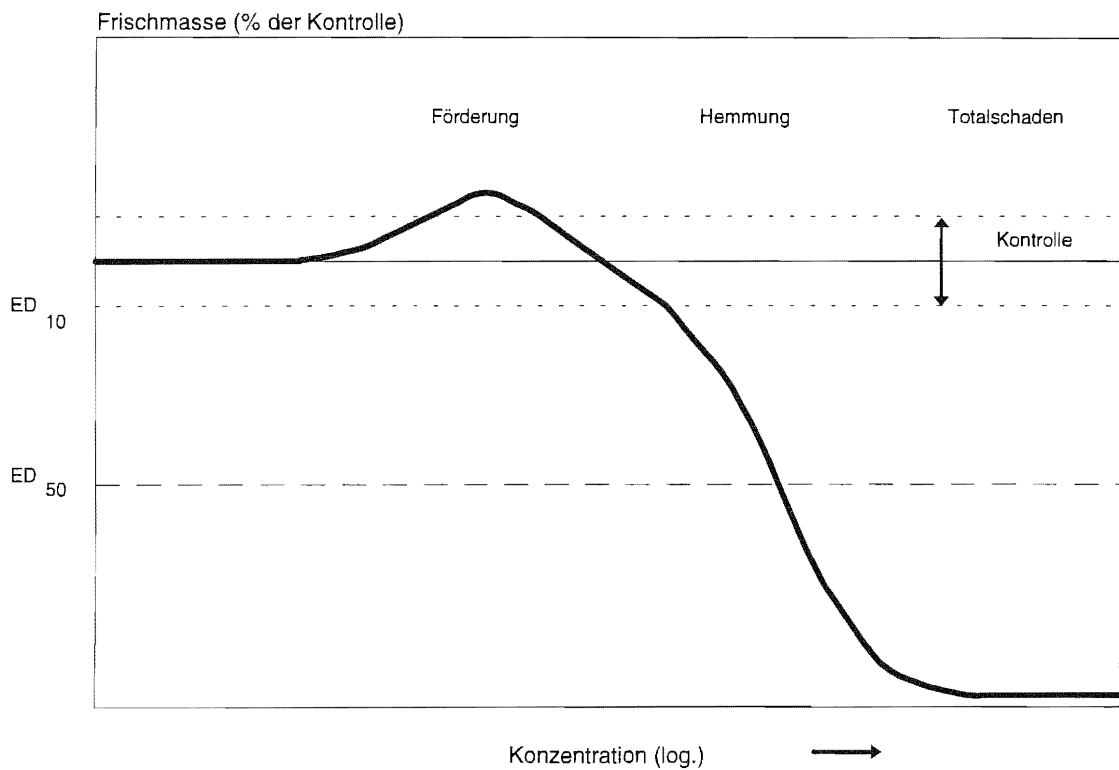


Abb. 5: Allgemeine Form der Dosis-Wirkungs-Beziehung zwischen Testpflanzen und phytotoxischen Stoffen

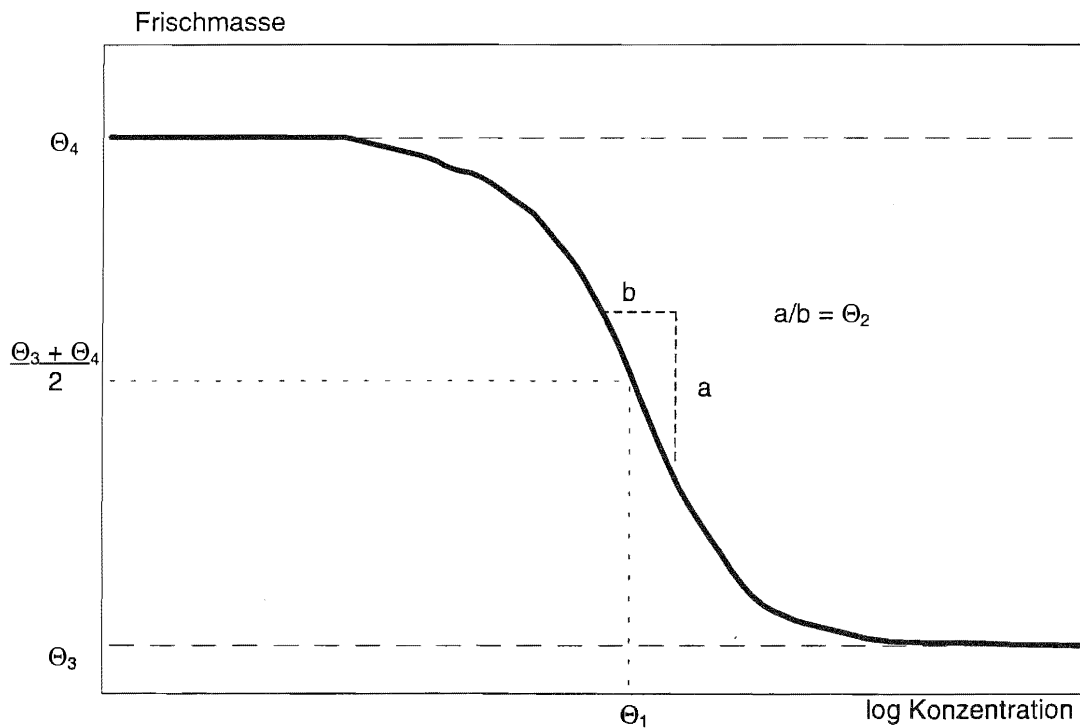


Abb. 6: Mathematische Form der Dosis-Wirkungs-Kurve nach Günther et al. (1989)

Die theoretischen Grundlagen der logistischen Kurvenanpassung mit dem Programm LOGFIT werden im folgenden kurz beschrieben. Die nichtlineare Regression beruht auf dem Prinzip der Minimierung der Abweichung der Meßwerte von der hineingelegten Funktion. Die Kurve (Abbildung 6) wird durch vier Parameter (Θ_{1-4}) bestimmt.

$$FM = \Theta_3 + \frac{\Theta_4 - \Theta_3}{1 + e^{-(\Theta_1 - \Theta_2 \cdot \ln \text{Konz})}}$$

mathematische Form der Dosis-Wirkungs-Kurve nach GÜNTHER et al. (1989)

Erklärung der Parameter:

- FM*: Frischmasse
 Θ_1 : Lageparameter, gibt ungefähr die Lage der ED_{50} auf der x-Achse an
 Θ_2 : Steigungsparameter, gibt an, wie steil der lineare Teil im mittleren Bereich verläuft
 Θ_3 : unteres Niveau der Kurve (Minimalgewicht, das bei hohen Konzentrationen erreicht wird)
 Θ_4 : oberes Niveau der Kurve, Gewicht der unbehandelten Pflanzen

Da in dieser Formel die Konzentration nicht gleich Null werden darf, wird im Programm für die Kontrolle automatisch eine Konzentration von 10^{-10} eingesetzt. Die Berechnung der vier Parameter für die jeweilige Funktion übernimmt das Programm. Für den Anwender des Programms sind Lage- und Steigungsparameter nicht von Bedeutung.

Das untere Niveau muß jedoch durch den Anwender festgelegt werden, da die Kurve sonst in den meisten Fällen in den negativen Bereich absinkt. Das obere Niveau kann aus zwei Gründen auf den Mittelwert der Kontrolle festgelegt werden:

1. Es wird sonst keine Konvergenz erreicht, d.h. die ED-Werte können nicht berechnet werden.
2. Der vom Programm berechnete Wert weicht zu stark (mehr als 10 % nach oben oder 1 % nach unten) vom tatsächlichen Mittelwert der Kontrolle ab.

Für jede Konzentration werden vom Programm Mittelwert und Standardabweichung berechnet. Die wichtigsten ED-Werte werden am Schluß der Berechnungen mit ihren Vertrauensgrenzen ausgegeben, um auch die Genauigkeit des Versuchs abzuschätzen.

Andere ökotoxikologische Tests

In den anderen ökotoxikologischen Tests mit quantitativen Merkmalen ist die Parametrisierung von Dosis-Wirkungs-Kurven bislang noch unüblich. Um die Auswertung zu veranschaulichen und um Hinweise auf die grundsätzliche Eignung zu erhalten, wurden Modellrechnungen mit realen Datensätzen von verschiedenen Tests durchgeführt. Dabei ist allerdings zu berücksichtigen, daß das Testdesign in der Regel auf die ANOVA-Auswertung zugeschnitten ist (wenige Dosisstufen mit verhältnismäßig großem Umfang). Die Modellierung von Dosis-Wirkungs-Kurven ist deshalb nicht für diese Art von Daten vorgesehen.

Regenwürmer (*C. Kula*)

Der Reproduktionstest mit Regenwürmern wird in der Regel mit 2 Dosisstufen durchgeführt. Mit diesen Daten ist die Ermittlung einer Dosis-Wirkungskurve nicht möglich. Es ist jedoch vorgesehen (ISO-draft 11268-2), für den Versuch mindestens drei Dosisstufen vorzuschreiben. In dem Versuch (mit 4 Wiederholungen/Dosis und 10 Tieren/Wiederholung) werden Reproduktion, Gewicht und Mortalität der Adulten ermittelt (siehe Kapitel 3, Tabelle 2). Die Richtlinie beinhaltet 4 verschiedene Validitätskriterien für die Kontrolle, bezogen auf Mortalität, Gewicht, Reproduktion und Variationskoeffizient der Jungtierzahlen.

Anläßlich eines internationalen Ringversuchs wurde der Reproduktionstest nicht mit nur zwei Dosisstufen, sondern mit 5 bis 6 Dosierungen durchgeführt. Die Versuche mit der Bezeichnung D1 und E1 sowie D2 und E2 hatten jeweils dieselbe Kontrolle. Diese Daten wurden parallel einer ANOVA zur Bestimmung der LOEC und NOEC und einer ECx-Auswertung unterzogen (Tabelle 3). Für die ECx-Auswertung wurde mit Hilfe der nicht-linearen Regression eine S-förmige Kurve modelliert (Bruce und Versteeg 1992; SAS-Statistik-Paket). Als x wurde hier 10 und 20 gewählt. Von den vier Versuchen sind die Versuche E2 in Abbildung 6 und der Versuch D1 in Abbildung 7 dargestellt. Die Streuung der Einzelwerte ist in beiden Fällen vom Mittelwert abhängig; im Versuch E2 passen die Daten nicht so gut auf den Kurvenverlauf.

Tabelle 3: Regenwurm-Reproduktionstest mit 4 verschiedenen Datensätzen, Auswertung mit Dunnett-Test und EC10 und EC20

Versuch	Dosis (mg/kg)	NOEC	LOEC	EC10	EC20
D 1	0.46/0.91/4.56/2 2.8/45.6/68.4	0.91	4.56	0.33 (0.14-0.77)	0.56 (0.27-1.13)
D 2	0.46/0.91/4.56/2 2.8/45.6	0.46	0.91	3.5 (2.4-4.9)	4.2 (3.1-5.7)
E 1	0.87/4.37/8.74/4 3.68/174.72	43.68	174.7	14.5 (3.7-56.9)	23.4 (8.0-70.0)
E 2	8.74/43.68/87.4/ 174.2/349.4	43.68	87.4	46.7 (20.7-105.0)	66.1 (33.8-127.4)

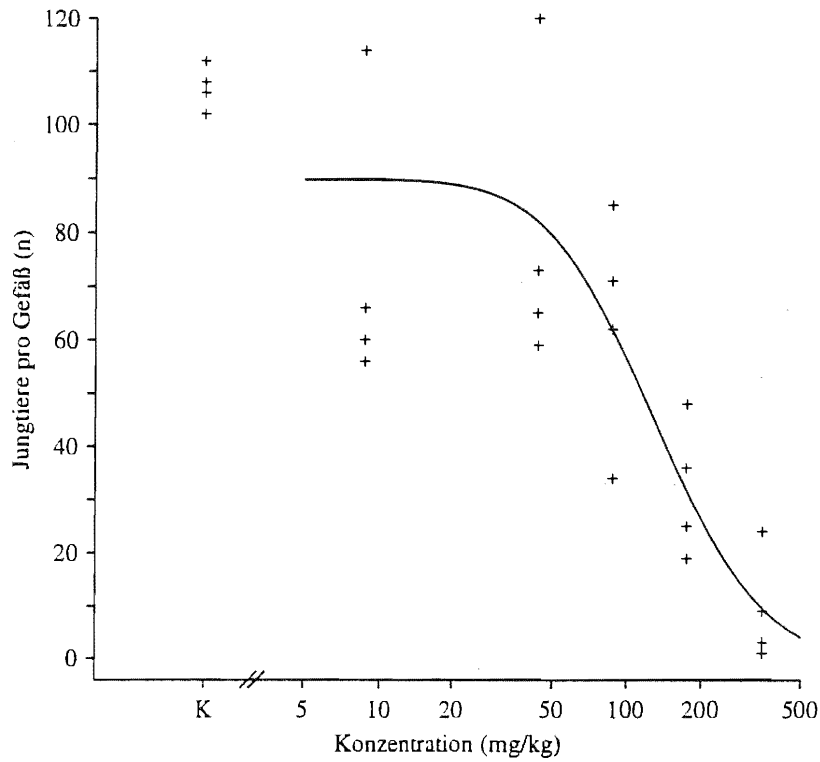


Abbildung 7: Darstellung der Dosis-Wirkungsbeziehung zu Datensatz E 2 aus Tabelle 3

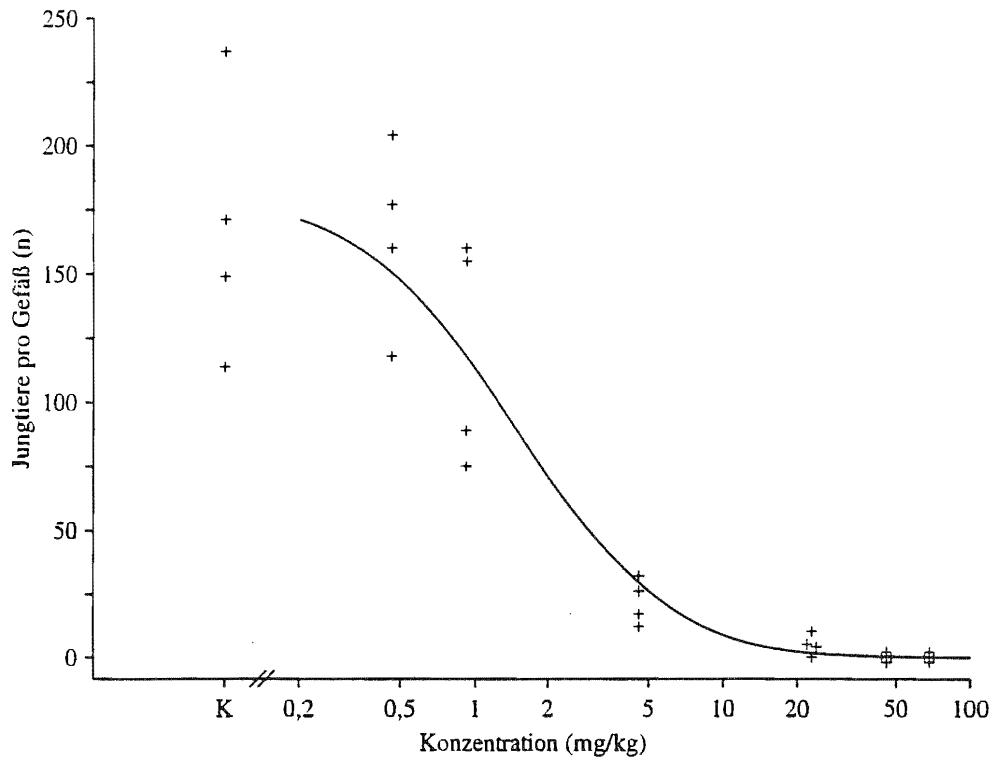


Abbildung 8: Darstellung der Dosis-Wirkungsbeziehung zu Datensatz D 1 aus Tabelle 3

Gewässerorganismen (M. Streløke)

Die in Tabelle 4 zusammengefaßten Daten stammen aus mehr oder weniger zufällig ausgewählten Studien, die im Zulassungsverfahren für Pflanzenschutzmittel eingereicht wurden und nach den entsprechenden OECD-Richtlinien 201, 202 (Teil II) und 204 durchgeführt wurden.

Tabelle 4: Langfristige/chronische Untersuchungen an Gewässerorganismen. Vergleich von ANOVA- und EC_x-Auswertung. Die Angaben in Klammern geben die jeweilige Reduktion (LOEC) bzw. das Konfidenzintervall (EC₂₀) an.

Versuch	Test-Konzentrationen	Endpunkt	NOEC	LOEC	EC ₁₀	EC ₂₀
Alge	0.1/0.3/0.6/1.0/ 1.8/3.2/5.6/10.0/ 32.0 mg/l	Zellzahl	3.2	5.6 (42%)	2.45 (1.6-3.7)	3.55 (2.6- 9.9)
Daphnie 1	0.04/0.16/0.62/ 2.6/ 10.3/42.3 µg/l	Jungtiere/ Weibch.	2.6	10.3 (34%)	0.82 (0.14- 4.74)	3.55 (1.1-11.4)
Daphnie 2	0.3/1.6/8.0/40.0/ 200.0/1000.0 µg/l	Jungtiere/ Weibch.	200	1000 (32%)	363 (30-4375)	630 (172-2290)
Fisch	0.2/0.5/1.1/2.3/ 5.0 µg/l	Wachstum	0.2	0.5 (8%)	2.0 (0.6-6.6)	18.2 (2.2-148)
		Mortalität	2.3	5.0 (60%)		2.5

Die EC₂₀ wurde hier als Ersatz für die NOEC verwendet. Die Auswertung zeigt allerdings, daß dieser Wert zu hoch ist, so daß die EC₁₅ oder EC₁₀ eher als geeigneter erscheinen.

Grundsätzlich zeigen die Auswertungen, daß keine entscheidenden Unterschiede zwischen den beiden Verfahren der Auswertung im Hinblick auf die Risikobewertung im Zulassungsverfahren bestehen. Allerdings ist die Übereinstimmung besser, wenn die Konzentrationsabstufung möglichst eng ist. Auffällig ist der große Unterschied zwischen dem auf das Wachstum der Fische bezogenen NOEC- und EC₂₀-Wert. In Abb. 9 ist der Verlauf der Kurve dargestellt. Offensichtlich führt die EC_x-Abschätzung zu einer Unterschätzung des Risikos, wenn die Dosis/Wirkungskurve sehr flach verläuft bzw. die Dosis/Wirkungs-Beziehung nicht klar erkennbar ist.

Prinzipiell ist für die EC_x-Auswertung das Vorhandensein einer Wiederholung pro Konzentration ausreichend. Demzufolge ist die Feststellung, daß - wie oben gezeigt - im Endergebnis keine deutlichen Unterschiede zwischen den Verfahren der Auswertung bestehen, nicht als Plädoyer für die ANOVA-Auswertung zu verstehen.

Sollte sich das obige Ergebnis in Untersuchungen bestätigen, in denen Testsysteme mit einer für ANOVA-Auswertungen ausreichenden Anzahl von Wiederholungen mit solchen verglichen werden, in denen nur ein Replikat geprüft wird, ist der EC_x-Ansatz zu bevorzugen, da weniger Testorganismen benötigt werden. Hierzu liegen aber zu wenig Daten vor. Derartige Untersuchungen müssen mit unterschiedlichen Testsystemen und -organismen durchgeführt werden, da die Streuung der Ergebnisse stark von diesen Variablen abhängen dürfte.

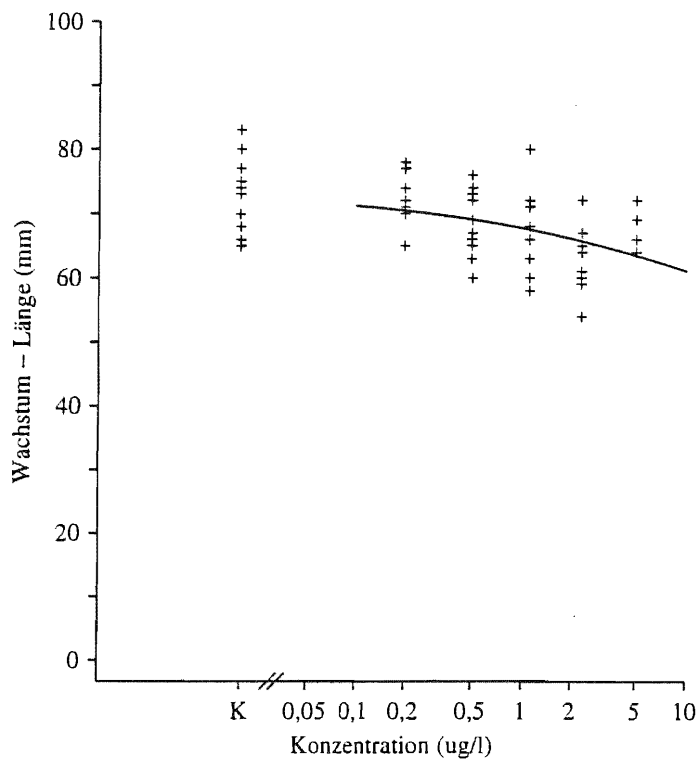


Abbildung 9: Darstellung der Dosis-Wirkungsbeziehung zum Beispiel „Fisch“ aus Tabelle 4

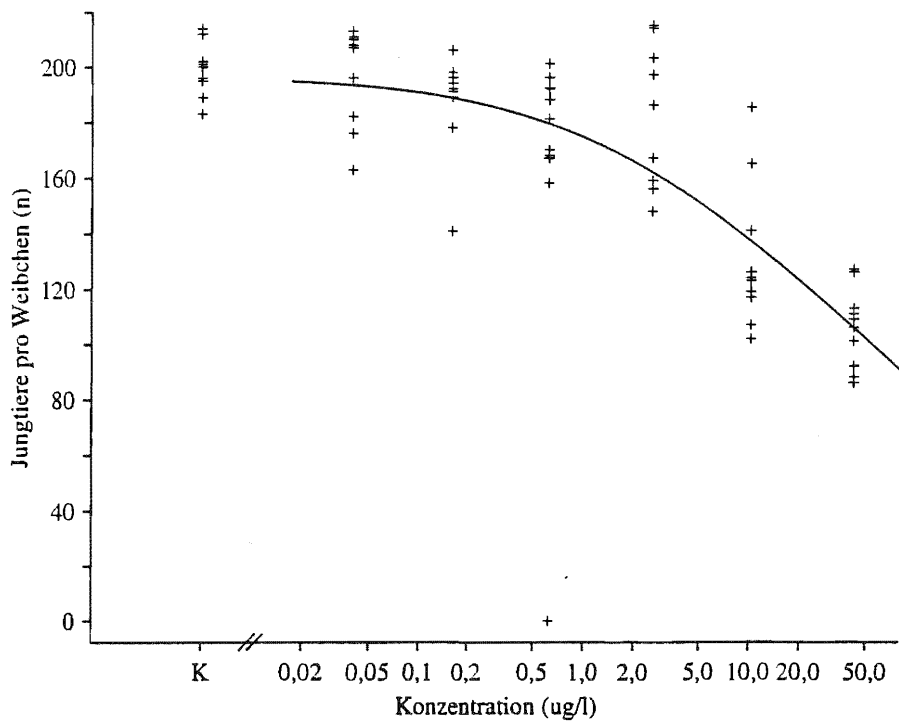


Abbildung 10: Darstellung der Dosis-Wirkungsbeziehung zum Beispiel „Daphnie 1“ aus Tabelle 4

Vögel (G. Joermann)

Der Reproduktionstest mit Vögeln wird stets nur mit drei Dosisstufen durchgeführt, so daß eine Parametrisierung der Dosis-Wirkungs-Kurve mit den vorhandenen Daten nicht möglich ist. Ein besonderes Problem besteht bei diesem Test darin, daß etwa ein Dutzend verschiedener Merkmale erhoben werden, von denen immer nur wenige von der Prüfsubstanz beeinflußt werden. Bei der derzeitigen Dosiswahl, die sich grob an der zu erwartenden Exposition orientiert, ist auch bei dem am stärksten beeinflußten Merkmal die Stärke des Effekts verhältnismäßig gering. In ca. 40 % der Versuche weicht in der oberen Dosisstufe überhaupt kein Merkmal von der Kontrolle ab. Einer generellen Anhebung der höchsten getesteten Dosis sind aus Tier-schutzgründen Grenzen gesetzt.

6.4 Diskussion

Bewertung der ECx-Methode

Grundsätzlich sahen die Teilnehmer des Fachgesprächs die Vorteile gegenüber der ANOVA-Methode, wobei allerdings die Zustimmung unterschiedlich stark war. Einigkeit bestand darin, daß das Konzept nicht pauschal und unkritisch zur Standardauswertung erklärt werden kann, sondern jeder Test einzeln hinsichtlich der Auswertemöglichkeiten geprüft werden muß. Bei der Entwicklung neuer und der Überarbeitung bestehender Richtlinien sollte in jedem Fall die Möglichkeit der EC-Punkt-Schätzung geprüft und ggf. in den Ringversuch mit einbezogen werden. Bei gleicher Eignung sollte nach Meinung der großen Mehrheit der Teilnehmer diese Auswertung der ANOVA-Methode vorgezogen werden. Einigkeit bestand darin, daß die Frage des Auswertungskonzeptes nur international behandelt werden kann.

Wahl des Modells

Eine Schwierigkeit bei EC-Punkt-Schätzungen besteht darin, das am besten geeignete mathematische Modell auszuwählen. Während bei quantalen Daten aufgrund langjähriger Erfahrung Probitanalysen (bzw. nahe verwandte davon) zum Standard geworden sind, hat sich für metrische Daten bisher kein einheitliches Modell durchsetzen können. Es gibt verschiedene Vorschläge; einige mit, andere ohne Reaktionsschwellen. Teilweise werden auch Besonderheiten integriert, wie fördernde Effekte bei niedrigen Konzentrationen (siehe 6.3, Abb. 5). Da bei längerfristigen ökotoxikologischen Versuchen die Meßgrößen und die Mechanismen, die den beobachteten Effekten zugrunde liegen, sehr vielfältig sind, gibt es wahrscheinlich kein Modell, das generell auf alle metrischen Daten paßt. In der Diskussion zeigte sich, daß die Ökotoxikologen hier noch erheblichen Beratungsbedarf durch Statistiker haben.

Festlegung eines geeigneten „x“

Eine Dosis-Wirkungs-Kurve wird durch zwei oder mehrere Parameter beschrieben, z. B. Lage- und Steigungsparameter. Daraus muß nun eine Kenngröße abgeleitet werden, die sich sinnvollerweise für die Einstufung der Prüfsubstanz bzw. für Risikoabschätzungen verwenden läßt. Da die meisten Modelle keine Wirkungsschwelle beinhalten (die Wirkung nähert sich mit abnehmender Dosis asymptotisch an Null), ist die Kenngröße eine Dosis, bei der ein bestimmter Effekt (x %) zu erwarten ist. Aus Sicht der Statistik sollte das "x" nicht zu klein sein, weil die Modellierung im Bereich sehr kleiner Effekte unsicher ist. Das "x" darf andererseits aber auch nicht zu groß sein, weil es sich sonst nicht mehr für Bewertungen eignet. In der internationalen Diskussion sind Werte zwischen EC05 und EC20. Die Diskussion ergab, daß eine Festlegung

ohne Kenntnis der Verhältnisse in den einzelnen Versuchen nicht möglich ist. Zu berücksichtigen sei die Variabilität des betrachteten Merkmals. Wichtig sei vor allem auch die biologische Bedeutung eines bestimmten Effekts bei dem jeweiligen Merkmal.

Berücksichtigung von ECx-Werten bei Bewertungen

Die international üblichen Verfahren der Risikoabschätzungen für Pflanzenschutzmittel und Chemikalien benutzen als Kenngrößen meist Expositions-Toxizitäts-Verhältnisse, wobei hinsichtlich der Toxizität ausdrücklich NOEC-Werte vorgesehen sind. ECx-Werte sind mit diesen Schemata nicht ohne weiteres kompatibel. Das bedeutet, daß eine Umstellung bei den Tests überhaupt nur dann möglich ist, wenn auch die Bewertungen angepaßt werden. Dies kann nur gelingen, wenn die Gremien für die Richtlinienarbeit und die für Bewertungen kooperieren. Eine Überlegung war in diesem Zusammenhang, das „x“ nach Auswertung vorhandener Daten so festzulegen, daß die Kenngröße ECx möglichst der NOEC äquivalent ist. Dadurch würde der Paradigmenwechsel erleichtert. Dagegen wurde aber argumentiert, daß dies das ECx-Konzept verwässert und zu Mißverständnissen bei der Interpretation der Werte führt.

Bei Verwendung von ECx-Werten bei Bewertungen ergibt sich das grundsätzliche Problem, daß ein ECx-Wert nicht als Dosis ohne Wirkung aufgefaßt werden kann, so daß ein zusätzlicher Sicherheitsfaktor auf den ECx-Wert angewendet werden müßte. Ein solcher Sicherheitsfaktor müßte per Konvention festgelegt werden, deshalb war die Meinung dazu mehrheitlich ablehnend. Es wurde vorgeschlagen, besser das „x“ so niedrig anzusetzen, daß man die ECx als „Dosis quasi ohne Wirkung“ oder, bei bestimmten Merkmalen, als „Dosis ohne schädliche Wirkung“ interpretieren kann. Eine andere Möglichkeit ist, die untere Vertrauensgrenze von ECx („benchmark“) als Kenngröße für die Bewertung heranzuziehen. Dies hätte den positiven Effekt, daß ein qualitativ guter Versuch (großer Versuchsumfang, niedrige Streuung) „belohnt“ wird. Diese Diskussionen sind auf internationaler Ebene noch nicht abgeschlossen.

7. Schlußfolgerungen und Empfehlungen

- Viele Richtlinien enthalten unzureichende Vorgaben für die statistische Auswertung. Bei der Entwicklung neuer Richtlinien sollte dieser Frage mehr Aufmerksamkeit gewidmet werden.
- In die Richtlinien sollten geeignete Validitätskriterien aufgenommen werden (z. B. zulässige Mortalität in der Kontrolle, zulässige Variabilität bestimmter Merkmale).
- Neue Richtlinien sollten in Ringtests validiert werden.
- Die ECx-Schätzung hat gegenüber der NOEC-Bestimmung prinzipielle Vorteile und sollte deshalb bevorzugt werden.
- Die ECx-Schätzung kann aber nicht pauschal zur Standardauswertung erklärt werden; jeder Test ist hinsichtlich der Auswertemöglichkeiten einzeln zu betrachten.
- Bei quantitativen Meßgrößen sind zahlreiche Modelle in der Diskussion, mit denen sich Dosis-Wirkungs-Funktionen parametrisieren lassen. Die Ökotoxikologen haben hier Beratungsbedarf durch Statistiker.

- ECx-Schätzungen sind nicht ohne weiteres mit den üblichen Bewertungsschemata kompatibel. Das bedingt eine gewisse Zähigkeit des Systems. Änderungen sind nur möglich, wenn die Entwickler von Richtlinien mit den Bewertungsgremien kooperieren.
- Prüfung und Bewertung von Pflanzenschutzmitteln und Chemikalien sind ein internationales Geschäft. Deshalb können Vorschläge für Verbesserungen und neue Konzepte nur im internationalen Rahmen umgesetzt werden.

8. Bibliographie

- Bruce, R.D., Versteeg, D.J. (1992): A statistical procedure for modeling continuous toxicity data. - *Environmental Toxicol. Chem.* 11, 1485-1494.
- Chapman, P.M., Caldwell, R.S., Chapman, P.F. (1996): A warning: NOECs are inappropriate for regulatory use. - *Environm. Toxicol. Chem.* 15, 77-79.
- Cousens, R., Marshall, C. (1987): Dangers in testing statistical hypotheses. - *Ann. Appl. Biol.* 111, 469-476.
- Collins, B.T. (1994): The avian reproduction study: distribution of the control data and statistical power analysis. Technical report no. 214, Canadian Wildlife Service, Headquarters.
- DIN (in Vorbereitung): Allgemeine Hinweise zur Planung, Durchführung und Auswertung biologischer Testverfahren (L1) DIN 38412 Teil 1.
- Hoekstra, J.A., Ewijk, P.H. (1993): Alternatives for the No-Observed-Effect Level. - *Environm. Toxicol. Chem.* 12, 187-194.
- Günther, P., Rahman, A., Pestemer, W. (1988): Quantitative bioassay for determining residues and availability to plants of sulfonylurea herbicides. - *Weed Research* 29, 141-146.
- MacLeod, D.A. (1994): Statistical analysis methods for avian reproduction experiments. Technical Report No. 211, Canadian Wildlife Service, Headquarters.
- Meister, R. (1994): Biometrical basics and principles of quantitative risk assessment based on animal experiments - models, benchmarks or NOELs for reproductive toxicity evaluation. *Informatik, Biometrie und Epidemiologie in Medizin und Biologie* 25 (4), 275-282.
- Noppert, F., van der Hoeven, N., Leopold, A. (1994): How to measure no effect. Workshop Report The Hague, September 9th, 1994.
- Pack, S., (1993): A review of statistical data analysis and experimental design in OECD aquatic toxicology test guidelines.
- Sachs, L. (1969): *Statistische Auswertungsmethoden*. Springer Verlag Berlin, S.279.
- Skalski, J.R. (1981): Statistical Inconsistencies in the Use of No-Observed-Effect Levels in Toxicity Testing. - *Aquatic Toxicology and Hazard Assessment: 4th Conference*. ASTM STP 737, D.R. Branson and K.L. Dickson (Eds), American Society for Testing and Materials, pp 377-387.

Anhang I

Tagesordnung

1. Begrüßung

2. Aktivitäten im Richtlinienprogramm der OECD (H. Köpp)

3. Versuchsanlage und Auswertung bei ökotoxikologischen Tests

Einführung: Übersicht über den gegenwärtigen Stand (C. Kula)

4. Möglichkeiten und Grenzen der Auswertung mit den derzeitigen statistischen Standardverfahren

- Pflanzenwachstumstest (W. Pestemer)
- Reproduktionstest mit Vögeln (G. Joermann)
- Versuche zur Auswirkung von Pflanzenschutzmitteln auf Laufkäfer und Spinnen (U. Heimbach)
- Versuche mit Regenwürmern (C. Kula)

5. NOEC versus EC-Punkt

Einführung: Gegenüberstellung der beiden Ansätze (H.T. Ratte)

- Versuchsdesign bei EC-Punkt-Schätzungen
- Robustheit der Verfahren (das Risiko, nicht auswertbare Datensätze zu erhalten)

6. EC-Punkt-Schätzungen: Eignung für bestimmte Tests und praktische Erfahrungen

- Reproduktionstest mit Vögeln (G. Joermann)
- Reproduktionstest mit Regenwürmern (C. Kula)
- Aquatische Ökotoxikologie (M. Streloke)

7. Kenngrößen bei EC-Punkt-Schätzungen

- Welches x ist aus statistischer Sicht geeignet?
- Welches x ist aus biologischer Sicht geeignet?
- Wie sollte der Vertrauensbereich berücksichtigt werden?
- Welche Kenngröße ist der NOEC äquivalent?

8. EC-Punkt-Schätzungen bei der Bewertung

- Definition von TER-Werten (Pflanzenschutzmittel, Chemikalien)

9. Verschiedenes

Anhang II

Teilnehmerverzeichnis

Dr. H. T. Ratte
Rheinisch-Westfälische
Technische Hochschule Aachen
Lehrstuhl für Biologie V
Worringerweg 1
52056 Aachen

S. Martin
Umweltbundesamt
IV 2.4
Postfach 330022
14193 Berlin

Dr. U. Klaschka
Umweltbundesamt
I 4.1
Postfach 330022
14193 Berlin

Dr. G. Joermann, R. Forster, Dr. C. Kula, Dr. M. Streloke
Fachgruppe Biologische Mittelprüfung
Abteilung für Pflanzenschutzmittel und
Anwendungstechnik
Biologische Bundesanstalt
Messeweg 11/12
38104 Braunschweig

R. Spangenberg
Fachgruppe Biologische Mittelprüfung
Abteilung für Pflanzenschutzmittel und
Anwendungstechnik
Außenstelle Kleinmachnow
Biologische Bundesanstalt
Stahnsdorfer Damm 81
14532 Kleinmachnow

H. Köpp
Koordinierungsgruppe
Abteilung für Pflanzenschutzmittel und
Anwendungstechnik
Biologische Bundesanstalt
Messeweg 11/12
38104 Braunschweig

A. Hillmer
DV-Gruppe
Abteilung für Pflanzenschutzmittel und
Anwendungstechnik
Biologische Bundesanstalt
Messeweg 11/12
38104 Braunschweig

Dr. E. Moll
DV-Gruppe
Abteilung für Pflanzenschutzmittel und
Anwendungstechnik
Biologische Bundesanstalt
Außenstelle Kleinmachnow
Stahnsdorfer Damm 81
14532 Kleinmachnow

Dr. U. Heimbach, K. Metge
Institut für Pflanzenschutz in Ackerbau und Grünland
Biologische Bundesanstalt
Messeweg 11/12
38104 Braunschweig

Dr. W. Pestemer
Institut für Ökologische Chemie
Biologische Bundesanstalt
Königin-Luise-Str. 19
14195 Berlin

Dr. H.P. Malkomes, Dr. P. Zwerger
Institut für Unkrautforschung
Biologische Bundesanstalt
Messeweg 11/12
38104 Braunschweig

Dr. Mueller
Institut für Ökotoxikologie
im Pflanzenschutz
Biologische Bundesanstalt Kleinmachnow
Stahnsdorfer Damm 81
14532 Kleinmachnow

Die **Berichte** aus der Biologischen Bundesanstalt für Land- und Forstwirtschaft erscheinen seit 1995 in zwangloser Folge.

Bisher erschienene **Berichte**:

- Heft 1, 1995 Sachverständigengutachten zur Genehmigung von Weihnachtsbaumkulturen (in Landschaftsschutzgebieten) unter Berücksichtigung von Herbizideinsätzen bzw. mechanischen oder kulturtechnischen Verfahren zur Unkrautbekämpfung und deren Folgewirkungen auf den Naturhaushalt. Dr. Gerd Heidler 100 S.
- Heft 2, 1995: Liste der zugelassenen Pflanzenschutzmittel (Stand: 1. Januar 1995). Bearbeitet von Dr. Achim Holzmann und Andreas Spinti, 63 S.
- Heft 3, 1995: Rechtliche Regelungen der Europäischen Union zur Prüfung und Zulassung von Pflanzenschutzmitteln und Wirkstoffen (Richtlinien, Verordnungen, Entscheidungen und Protokolle) (Stand: 1. Juni 1995). Bearbeitet von Dr. Jörg-Rainer Lunde, 233 S.
- Heft 4, 1995: Verzeichnis der Wirkstoffe in zugelassenen Pflanzenschutzmitteln (ehemals Merkblatt Nr. 20), (Stand: November 1994). Bearbeitet von Dr. Günter Hoffmann, 86 S.
- Heft 5, 1995: Spritz- und Sprüngeräte für Flächenkulturen
Auszug aus der BESCHREIBENDEN PFLANZENSCHUTZLISTE
-Teil Geräte-.
Bearbeitet von Dr.-Ing. Heinz Ganzelmeier, Sabine Gebauer, Hans-Joachim Wehmann und Siegfried Rietz, 170 S.
- Heft 6, 1995: Information Exchange and Prior Informed Consent (PIC) Procedure in the Export and Import of Pesticides in the Framework of the FAO Code of Conduct. Bearbeitet von Dr. Achim Holzmann, 111 S.
- Heft 7, 1995: Workshop Integrated Pest Management
November 2nd 1995, Kleinmachnow.
Bearbeitet von Dr. Holger Beer, 39 S.
- Heft 8, 1995: Art und Menge der in der Bundesrepublik Deutschland abgegebenen und der exportierten Wirkstoffe in Pflanzenschutzmitteln (1987-1994)
Ergebnisse aus dem Meldeverfahren nach § 19 des Pflanzenschutzgesetzes.
Bearbeitet von Dr. Hans-Hermann Schmidt, Dr. Achim Holzmann und Edelgard Adam, 65 S.
- Heft 9, 1995: Arbeitsschutz und Arbeitssicherheit im öffentlichen Dienst (Stand: Juni 1995). Dirk Altwein, 16 S.
- Heft 10, 1996: Zur Umsetzung biometrischer Verfahren in SAS mit Beispielen aus dem Pflanzenschutz.
Eckart Moll, 185 S.
- Heft 11, 1996: Liste der zugelassenen Pflanzenschutzmittel (Stand: 1. Januar 1996)
Bearbeitet von Dr. Achim Holzmann und Andreas Spinti, 63 S.
- Heft 12, 1996: Methodische Anleitung zur Bewertung der partiellen Resistenz und die SAS - Anwendung RESI. Eckart Moll, 60 S.
- Heft 13, 1996: Saatgutbehandlung von Getreide und Beschreibende Liste - Beizgeräte (Stand: Dezember 1995). Bearbeitet von Dr. Helmut Ehle, Dr. Günter Menschel, Dr. Wolfgang Radtke, Siegfried Rietz, Friedrich-Otto Ripke, 48 S.
- Heft 14, 1996: Die SAS-Anwendung FELD_VA-konstruktion des Lageplanes und der varianz-analytischen Auswertung ein- bis dreifaktorieller Feldversuche. Dr. Eckart Moll, 43 S.
- Heft 15, 1996: Dokumentation der Forschungsvorhaben - Forschungsaufgaben der BBA unter besonderer Berücksichtigung ihrer „Drittmittelforschung“ - laufende Vorhaben der BBA, Stand Januar 1996. Dr. Holger Beer Dr. Heinrich Brammeier, 145 S.
- Heft 16, 1996: Assessing Volatilization of Pesticides: A comparison of 18 Laboratory Methods and a Field Method. Bearbeitet von Ulrike Walter, Dr. Matthias Frost, Garnet Krasel, Prof. Dr. Wilfried Pestemer, 44 S.