

Vor- und Nachdenkliches ... zum wahren Erfolg des 3R Prinzips

Marcel Leist, Suzanne Kadereit und Stefan Schildknecht

Doerenkamp-Zbinden Lehrstuhl für in vitro Methoden zum Tierversuchersatz, naturwissenschaftliche und mathematische Fakultät, Universität Konstanz, Deutschland

Einleitung

Ist es möglich, den „Wert“ einer wissenschaftlichen Disziplin zu bemessen? Wie kann diese Disziplin dann in einem Gesamtrahmen eingeordnet werden, und nicht zuletzt, ist eine solche Überlegung überhaupt gerechtfertigt? Von einigen Seiten mögen bei solchen Fragen Zweifel aufkommen; andere werden die Forschung im Bereich der Alternativmethoden zu Tierversuchen als kategorischen Imperativ ohne jegliches wenn und aber betrachten. Schließlich gibt es noch eine etwas differenziertere Sichtweise auf diese Problematik, die sehr häufig von Personen in gesetzgebender Funktion, von privaten wie öffentlichen Geldgebern, Tierschutzbeauftragten, und nicht zuletzt von jungen Studierenden, die sich entschließen, ihre Karriere im Bereich der Alternativmethoden zu beginnen, vertreten wird. Diese unterschiedlichen Betrachter stellen früher oder später eine Kosten-Nutzen Rechnung auf und wägen verschiedene Aspekte gegeneinander ab. Mit diesem Ansatz stehen sie jedoch nicht alleine da: schließlich wird ein Großteil der Steuerzahler und Wähler ähnliche Gedanken anstellen. Aus diesem Grund erscheint es mehr als gerechtfertigt, einige grundsätzliche Gedanken über den bereits erwähnten „Wert“ des 3R Ansatzes (Replace, Reduce, Refine; Ersetzen, Reduzieren, Verbessern), wie er von Russell und Burch 1959, sowie in der Bologna-Erklärung 1999 aufgestellt wurde, zu formulieren. Nach rein qualitativen Gesichtspunkten wird die wachsende Bedeutung von Alternativmethoden schon durch die Einrichtung einer eigenen Institution (ECVAM), deren Aufgabe die Etablierung und Bewertung alternativer *in vitro* Methoden ist, durch die EU unterstrichen. Ferner stützt sich bereits ein Großteil der neueren EU-Gesetzgebung (REACH) auf die Verwendung von Alternativmethoden

(REACH, 2006). In diesem Zusammenhang arbeitet die EU-Kommission bereits seit einiger Zeit in enger Partnerschaft mit der Industrie (EPAA, 2007), und zusätzlich wurden an Universitäten neue Lehrstühle für die Entwicklung von Alternativmethoden für die Entwicklung von Alternativmethoden zu Tierversuchen eingerichtet (Leist, 2006; Wendel, 2002).

Eine Reihe quantitativer Versuche, die Entwicklung von Alternativmethoden zu erfassen, stützt sich als alleinige Bewertungsgrundlage auf den quantitativen Tierverbrauch innerhalb der EU oder in den einzelnen Mitgliedsstaaten. Ferner wurde auch die Anzahl der neuen OECD Richtlinien, die Alternativmethoden als Evaluationsmethode beinhalten, statistisch erfasst und zur Auswertung herangezogen. Basierend auf diesen Zahlen erscheint der Fortschritt in der Implementierung des 3R Prinzips relativ langsam und wenig bedeutsam. Es ist uns daher ein besonderes Anliegen an dieser Stelle klar hervorzuheben, dass durch eine solche Herangehensweise die wahren Erfolge und Fortschritte, die bereits im Bereich der Entwicklung von Alternativmethoden erzielt worden sind, bei weitem unterbewertet wurden. Aus diesem Grund möchten wir im Folgenden einige Denkfehler diskutieren, die regelmäßig in der öffentlichen Debatte auftreten und die in großem Maße zu der bereits erwähnten Unterbewertung des 3R Ansatzes geführt haben.

Denkfehler I: Zahlen der toxikologischen Tierversuche als einzige Bemessungsgrundlage

Das relativ klar umrissene und standardisierte Versuchsspektrum in der Toxikologie eignet sich besonders gut für die Anwendung von Alternativmethoden. Das relativ starke Engagement der 3R Forschung im Bereich der Sicherheitstestung

wird durch die teilweise hohen Belastungsgrade der in diesem Bereich durchgeführten Tierversuche leicht verständlich. Hierbei darf jedoch nicht übersehen werden, dass die besonders starke Unterstützung seitens der Industrie sowie der Regierungen die vorwiegende Triebkraft der Forschung auf diesem Gebiet darstellt – und diese in erster Linie auf wirtschaftlichen Interessen beruht. Einige Zahlen sollen einen besseren Überblick liefern: Der Gesamtverbrauch an Versuchstieren in der EU betrug 2002 ca. 11 Millionen (EU, 2005) und im Jahr 2005 ca. 12 Millionen Tiere (EU, 2007). Davon wurden nur rund 10% (2002), respektive 8% (2005), für toxikologische Studien herangezogen. Somit wird schnell klar, dass die Ermittlung der Fortschritte von Alternativmethoden allein auf der Grundlage statistischer Befunde im toxikologischen Bereich vergleichsweise eingeschränkt ist. Die Problematik dieses Bewertungsansatzes wird noch deutlicher, wenn bedacht wird, dass OECD-Richtlinien nur für verschiedene Unterklassen an toxikologischen Tests relevant sind, so z.B. für die Risikoabschätzung von Substanzen aus der chemischen Industrie, jedoch nicht für die von Medikamenten. Dies hat zur Folge, dass das alleinige Aufzählen bereits angewandter oder validierter Testverfahren zum Ersatz von Tierversuchs OECD-Richtlinien zu einer deutlichen Unterschätzung der wahren Entwicklung auf dem Gebiet der 3R Forschung führt (Gruber und Hartung, 2004). Tatsächlich wird mit diesem Vorgehen gerade einmal ein einstelliger Prozentsatz aller in Tierexperimenten verbrauchten Tiere (ca. 200.000 von 12 Millionen) überhaupt erfasst. Um eine bessere Vorstellung vom tatsächlichen Fortschritt des 3R Konzepts zu bekommen muss man das Augenmerk auch auf andere Bereiche in denen Tierversuche durchgeführt werden richten (Gruber und Hartung, 2004). Für Ausbildungs-

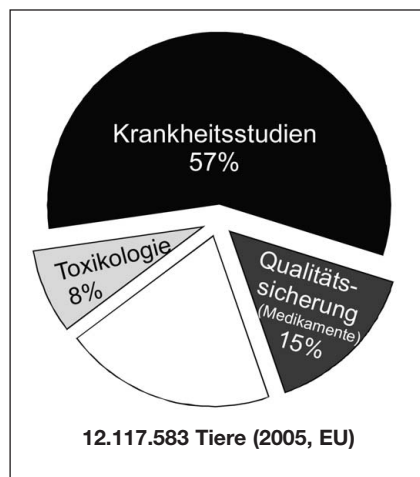


Abb. 1: Tierversuch in der Europäischen Union (EU) im Jahre 2005

zwecke beispielsweise ist der Verbrauch innerhalb des Zeitraums von 2002 bis 2005 um rund 50% zurückgegangen (von 3,2% auf 1,6%) (siehe Beispiele in Dewhurst 2006; Gruber and Dewhurst 2004). Ein anderer Bereich mit einem hohen Durchsatz an Versuchstieren (ca. 15% aller Versuchstiere in der EU = 1,7 Millionen/Jahr) stellt die Herstellung und Qualitätssicherung von Medikamenten dar. Als positive Beispiele seien hier zum einen neu etablierte *in vitro* Pyrogentests, sowie die Einführung von ELISA Tests zur Chargenkontrolle in der Impfstoffproduktion hervorzuheben (Hartung, 2001; Rosskopf-Streicher, 2001; Hendriksen, 2006; Montag et al., 2007; Hoffmann et al., 2005). Im Bereich mit dem höchsten Tierbedarf (> 50%; > 6 Millionen/Jahr), der Kategorie „*Study of Disease / Untersuchungen zu Krankheitsmechanismen*“, soll hier zum Beispiel auf das Verbot von Aszites-Mäusen und die Einführung der *in vitro* Produktion monoklonaler Antikörper verwiesen werden (Kuhlmann et al. 1989). Ferner konnten *in vitro* Systeme z.B. in der Borrelioseforschung (Kröber und Guerin, 2007), in der Angiogeneseforschung (Bahramsoltani, 2006), in der Erforschung genetischer Defekte (Akyüz und Wiesmüller, 2003; Kreja, 2003), oder in der Erforschung des Morbus Parkinson (Lotharius et al., 2005) erfolgreich eingeführt werden. Des Weiteren konnte eine Reihe von tierbasierten Nachweismethoden zur Bestimmung von Hormonspiegeln durch den Einsatz von diversen *in vitro* Testverfahren ersetzt werden. Eine

andere bemerkenswerte Ersatzmethode ist der „*in vitro colony forming assay*“ (Pessina et al., 2005; ECVAM DB-ALM, 2006), der zur Vorhersage von Myelosuppression durch Chemikalien, aber auch zur Untersuchung der zugrunde liegenden Mechanismen durchgeführt wird. Ein weiterer Test, für den jährlich hunderttausende Mäuse ihr Leben lassen, ist die Qualitätskontrolle von Speisemussheln. In diesem Zusammenhang sind besonders die positiven Bemühungen Deutschlands und Neuseelands hervorzuheben, die zum Ziel haben, den Maustest durch ein chemisch-analytisches Verfahren zum Nachweis von Muscheltoxinen zu ersetzen (Biosecurity, New Zealand, 2007). Auf gleicher Ebene sollte die Etablierung eines Fischembryo-Tests in der Wasserqualitätskontrolle genannt werden (Nagel, 2004), durch den der Einsatz ausgewachsener Fische überflüssig wurde.

Denkfehler II: Naiver Umgang mit statistischen Daten

In der EU werden alle drei Jahre Statistiken über den Tierversuch zu Forschungszwecken erstellt. Zusätzlich werden von den meisten Mitgliedsstaaten

jährliche Statistiken über den jeweiligen nationalen Verbrauch geführt. Diese Statistiken werden regelmäßig herangezogen wenn es darum geht, den Erfolg oder Misserfolg neu eingeführter Alternativmethoden zu bewerten, was jedoch dadurch, dass in den einzelnen Ländern unterschiedliche Bemessungsgrundlagen gelten, sehr schwer wenn nicht gar unmöglich ist. Beispielsweise wurden zwischen den Jahren 2002-2005 neue Mitgliedsstaaten in die EU aufgenommen. Daher wurde für das Jahr 2005 ein signifikant erhöhter Verbrauch an Versuchstieren in der EU vermeldet. Dies ist jedoch nicht auf ein Versagen in der Etablierung von Alternativmethoden zurückzuführen sondern liegt hauptsächlich an der Erhöhung der Bezugsgrundlage, in diesem Falle dem Anstieg der Anzahl von EU-Mitgliedsstaaten. Eine Gefahr der systematischen Fehlinterpretation bezüglich der Durchschlagkraft von 3R Methoden besteht auch bei Änderungen in der Definition eines Tierversuchs (z.B. Aufnahme/Nicht-Aufnahme der Organentnahme am getöteten Tier in die Kategorie Tierversuch, oder Mitzählung von Embryonen bei der Tierversuchstatistik).

Was sich jedoch als weitaus größeres Problem als diese erwähnten rein technischen Schwierigkeiten erweist, ist die weit

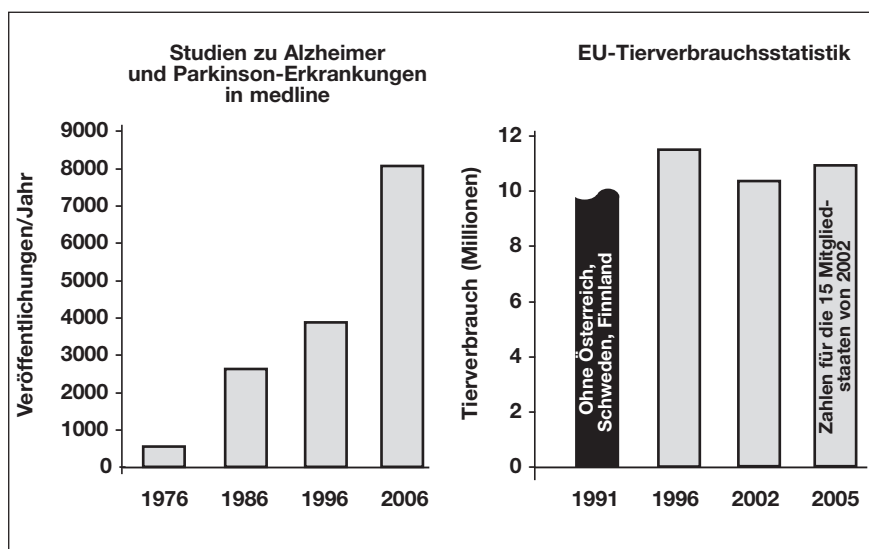


Abb. 2: Gleichbleibende Tierzahlen bei expandierender Wissenschaft

Die Anzahl Veröffentlichungen in fast allen Wissenschaftsgebieten steigt steil an (zum Beispiel die Publikationen zu Morbus Alzheimer und Parkinson, obere Graphik). Die Tierzahlen, die in der EU verbraucht wurden, sind in den letzten 15 Jahren relativ konstant geblieben (untere Graphik: der Balken für 2005 beinhaltet nur die Tierzahlen für die 15 Mitgliedsstaaten, die schon 2002 Teil der EU waren).

verbreitete alleinige Zugrundelegung tierstatistischer Daten zur Bewertung der Fortschritte und Erfolge des 3R Konzeptes. Angenommen wir hätten eine unveränderte Anzahl EU Mitgliedsländer sowie einheitliche statistische Auswerteverfahren: Würde unter diesen Umständen eine gleich bleibende Zahl an Tierversuchen ein Versagen des 3R Konzeptes bedeuten? Die Antwort ist ein klares NEIN. Das Forschungsvolumen im biomedizinischen Bereich sowie die Anzahl Wissenschaftler und Veröffentlichungen sind in den letzten Jahren sprunghaft angestiegen. In den vergangenen 25 Jahren wurden z.B. die Aufwendungen der Pharmaindustrie für Forschung und Entwicklung bis zu achtfach erhöht (DiMasi, 2003). Im gleichen Zeitraum hat sich die Anzahl Veröffentlichungen im biomedizinischen Sektor circa verzehnfacht. Als anschauliches Beispiel soll hier die Entwicklung der Anzahl von Veröffentlichungen im Bereich der Parkinson und Alzheimer Forschung dienen (Abb. 2). Wenn also innerhalb eines bestimmten Zeitraumes in der näheren Vergangenheit der Verbrauch an Versuchstieren auf gleichem Niveau geblieben ist, so kann dies zu Recht als großer Erfolg für den Einsatz von Alternativmethoden gewertet werden. Somit können in diesem Zusammenhang diejenigen Länder oder Firmen als erfolgreich angesehen werden denen es gelungen ist, den Versuchstierverbrauch im Verhältnis zu einem dramatisch gestiegenen Forschungsvolumen konstant zu halten. Es soll an dieser Stelle aber auch an das Phänomen der Globalisierung erinnert werden (Bottini et al., 2007). Statistiken über den Tierversuch könnten sehr leicht durch die Auslagerung von Tierversuchen in Nicht-EU Staaten geschönt werden. Solche Tendenzen dürfen auf keinen Fall ignoriert werden, da sie nicht im Sinne des 3R Gedankens und noch viel weniger im Sinne der betroffenen Versuchstiere sind.

Denkfehler III: Unterschätzung von 3R aufgrund der Messung von Veröffentlichungsfrequenzen

Aus der vorangegangenen Argumentation wird rasch klar, dass es einer weitaus differenzierteren Herangehensweise als der bloßen Betrachtung der Tierversuchsvolumens

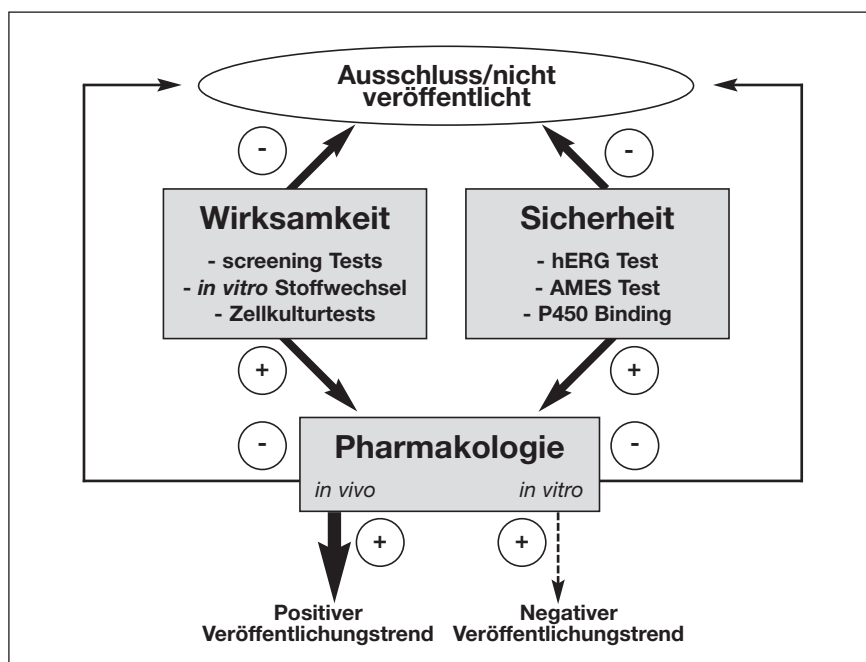


Abb. 3: Veröffentlichungen spiegeln die Nutzung von Alternativmethoden in der Medikamentenentwicklung nicht richtig wider

In den frühen Phasen der Medikamentenentwicklung werden viele *in vitro* Ansätze verwendet („Wirksamkeit“ und „Sicherheit“). Ungeeignete Verbindungen (Minuszeichen) werden ausgeschlossen und diese Daten werden meist nicht veröffentlicht. Geeignete Verbindungen (Pluszeichen) werden pharmakologisch weitergetestet und, falls sie sich als ungeeignet erweisen, danach ausgeschlossen. Bei Verbindungen, die weiter entwickelt werden, ist die Chance höher, dass die *in vivo* gewonnenen pharmakologischen Ergebnisse publiziert werden, als bei Ergebnissen, die mit Alternativmethoden gewonnen wurden.

zahlen bedarf, um Fortschritte in der Einführung des 3R Ansatzes aufzuzeigen. Eine häufig angewandte Methode im modernen Wissenschaftsbetrieb ist das Quantifizieren von relevanten Veröffentlichungen zu einem bestimmten Thema. Dieses Verfahren wird zweifellos auch für den 3R Bereich eine positive Entwicklung aufzeigen. Jedoch kann die tatsächliche Entwicklung des 3R Konzeptes allein durch die Erfassung aller relevanten Veröffentlichungen leicht unterschätzt werden. Als Beispiel hierzu soll in diesem Zusammenhang die Medikamentenentwicklung herangezogen werden. Mehr als 99% aller potentiellen Verbindungen werden bereits im Frühstadium dieses Prozesses ausgeschlossen. Die Daten, die zu solchen Substanzen gewonnen wurden, werden in aller Regel nicht veröffentlicht. Dies erklärt sich zum einen aus einer nachvollziehbaren Tendenz der beteiligten Unternehmen zur Geheimhaltung. Ferner besteht aber in der Veröffentlichungskultur auch ein starker Trend Negativ-Daten,

also Daten, bei denen kein Effekt einer Substanz auf das potentielle Ziel (target) zu beobachten war, einfach nicht zu publizieren. Ein nicht unerheblicher Teil dieser Art von Daten wurde jedoch mit Hilfe von Alternativmethoden gewonnen. Durch die Nichtveröffentlichung dieser Daten kommt es zur Unterrepräsentation der Alternativmethoden (Abb. 3). Zusätzlich eliminieren einfache sicherheitspharmakologische *in vitro* Screeningmethoden (wie der „human ether-a-gogo related receptor (HERG) Assay“ oder der Ames-Test) viele Substanzen aus der Medikamentenentwicklung über die dann niemals publiziert wird (Abb. 3). In den fortgeschrittenen Stadien der Medikamentenentwicklung, in denen vermehrt *in vivo* Tests eingesetzt werden, ergibt sich hingegen ein gegensätzliches Bild. So ist es immer noch üblich und möglich Negativ-Daten, die mit Hilfe von Tierversuchen gewonnen wurden, zu veröffentlichen, während dies mit *in vitro* Daten aus einem Alternativverfahren nahezu unmöglich ist.

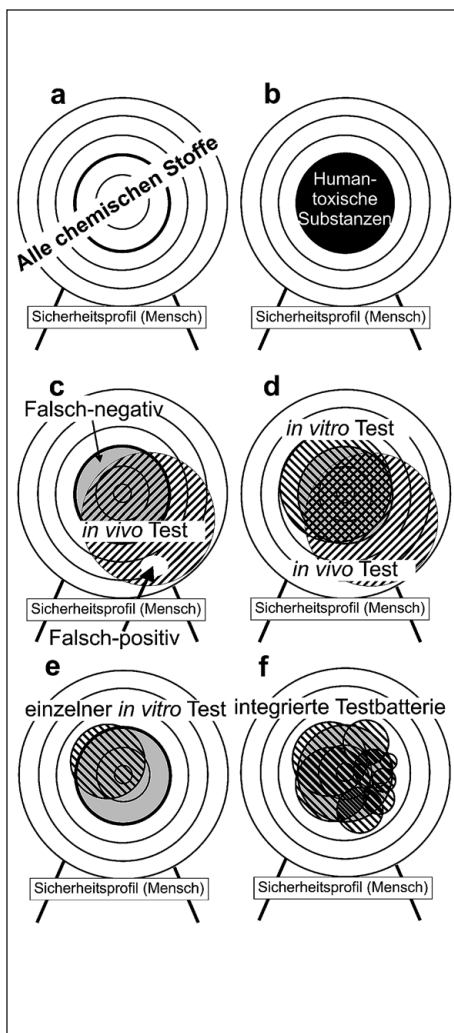


Abb. 4: Überlappung von Testergebnissen und der Realität bei der Sicherheitstesting

a) Alle chemischen Verbindungen, mit denen der Mensch in Kontakt kommt, sind als Zielscheibe dargestellt.

b) Die für den Menschen toxischen Verbindungen sind in der Mitte der Zielscheibe dargestellt (graues Feld). Gute Tests müssen ein hohes Maß an Überlappung mit diesem Bereich aufweisen.

c) Ein theoretisches Beispiel eines typischen *in vivo* Sicherheitstests ist gezeigt (schraffierte Fläche). Der Überlappungsbereich schließt die meisten toxischen Verbindungen ein, aber es gibt einige Falschnegative (oben links) und viele Falschpositive (unten rechts).

d) Ein Beispiel (z.B. Pyrogenität) eines *in vivo* und *in vitro* Tests für die gleiche Sicherheitsdomäne wird gezeigt. Hier hat der *in vitro* Test weniger Falschpositive und weniger Falschnegative als der *in vivo* Test. Trotz guter Überlappung mit der Realität überlappt der *in vitro* Test nur partiell mit der Fläche des *in vivo* Tests. Wenn der *in vivo* Test als Goldstandard angesehen würde, würde der *in vitro* Test durchfallen.

e) Zeigt ein theoretisches Beispiel einer Alternativmethode als Einzelelement einer integrierten Teststrategie (z.B. Spermatoxizität als Teil der Entwicklungstoxizität). Für sich alleine hat der Test schlechte Sicherheitscharakteristika (viele Falschnegative).

f) Zeigt den gleichen Test wie in e) als Teil einer integrierten Testbatterie mit optimalen Sicherheitscharakteristika (keine Falschnegativen, wenige Falsch-positive).

Das kann sehr leicht fälschlich zum Eindruck führen, dass *in vitro* Verfahren eine vermehrt untergeordnete Rolle spielen.

Denkfehler IV: Ersatz von *in vivo* durch *in vitro* im Verhältnis 1:1

Die klassische Toxikologie hat ein System der Sicherheitstesting entwickelt, das sich voll und ganz an den Möglichkeiten und Einschränkungen von Tiermodellen orientiert. Diese Art des Testverfahrens hat die gegenwärtigen Spielregeln für Sicherheitstestungen festgelegt – Spielregeln des „Tierversuchsspiels“. Diese Vorschriften sehen unter anderem eine Einteilung in bestimmte Gefahrentypen vor (z.B. Mutagenitätspotential, Reproduktionstoxizität,

usw.) und liefern eine Einschätzung über die zu verwendenden Dosen und deren Übertragbarkeit auf die Situation im Menschen. Der Versuch, gegen die Spielregeln des „Tierversuchsspiels“ zu verstoßen, wird leicht als „Versagen“ oder gar als „Hang zu unwissenschaftlichem Arbeiten“ gesehen. Um sich an diese Spielregeln zu halten, und um ja nicht von der etablierten Gefahrtypeneinteilung abzuweichen, hat der 3R Ansatz in der Vergangenheit oft den Versuch unternommen, etablierte *in vivo* Modelle im Verhältnis 1:1 durch Alternativmethoden zu ersetzen. Als erfolgreiches Beispiel sei der Phototoxizitätstest erwähnt der durch den Fibroblasten-Toxizitätstest ersetzt wurde (OECD Test Guideline 432, 2007). Betrachtet man jedoch andere Gefahrentypen, wie z.B. Reproduktionstoxizität oder Sensiti-

vierungspotential, erscheint eine 1:1 Substitution der gegenwärtig verwendeten Tiermodelle durch Alternativmethoden sehr unwahrscheinlich. Dennoch wurden viele Fortschritte in der Entwicklung von 3R Methoden in den zwei genannten Domänen erzielt. Diese Tests erfassen jedoch jeweils nur einzelne Schritte eines komplexen biologischen Ablaufs. Zum Beispiel wird aus dem Prozess der Sensitivierung die Bindung von Haptenen an Proteine oder die Aktivierung von dendritischen Zellen als Parameter verwendet, und aus dem Prozess der Entwicklungstoxizität die Bindung an Steroidrezeptoren, die Abtötung von Spermatozoen oder die Beeinträchtigung der Differenzierung embryonaler Stammzellen. Solche Einzelpunkte und -messgrößen müssen in integrierten Teststrategien zusammengesetzt werden welche dann letztlich alle Einzelschritte des gesamten biologischen Vorgangs abdecken sollen (Abb. 4). Keiner der Einzeltests wird mit dem ursprünglichen Tierversuch vergleichbar sein, kann aber innerhalb der Teststrategie, welche sich aus verschiedenen *in vitro* Methoden zusammensetzt oder aus *in vitro* Methoden und wenigen reduzierten oder verbesserten Tierversuchen besteht, einen entscheidenden Baustein darstellen (Combes, 2007). Somit führt die Erwartung einer 1:1 Substitution wieder zu einer starken Unterschätzung der Fortschritte der 3R Forschung. Weiter verschlimmert wird dies dadurch, dass diese Sichtweise die Entwicklung und Anwendung von 3R Methoden verhindert indem sie ihnen die Regeln des „Tierversuchsspiels“ auferlegt. So lange sich Alternativmethoden rigide an Tierexperimenten als Goldstandards messen müssen werden solche Methoden auch die inhärenten Limitationen des Tierversuchs beibehalten (Abb. 4) und völlig andere, innovative Ansätze haben keine Chance, dem Validierungsprozess standzuhalten (Hartung, 2007). Ironischerweise ist die Sachlage der Unterschätzung in den Gebieten, in denen der Tier-Goldstandard nicht so gut etabliert ist, noch weit ausgeprägter. Zum Beispiel gibt es die Sicherheitstesting von Biologika oder Nanopartikeln noch nicht lange und das Tierexperiment ist weniger gut standardisiert als für die klassischen Kleinmoleküle. Hier können Alternativmethoden schon jetzt wertvolle Informa-

tionen für die Sicherheitsprüfung liefern, aber da es keinen klaren Goldstandard gibt, den man ersetzen könnte, ist der Erfolg von 3R Methoden schwer messbar und wird deshalb in vielen klassischen statistischen Ansätzen, die auf quantitativen Zahlen beharren, unterschätzt.

Denkfehler V: Monodimensionale Fokussierung auf „Replacement“

Die Bemessung des Erfolgs der 3R Methoden allein aufgrund von ersetzten (1R) Tierversuchen oder OECD Richtlinien unterschätzt den Erfolg des integrierten 3R Ansatzes bezüglich der Minderung des Leidens der Tiere (Abb. 5). Obwohl es keinen Zweifel gibt, dass das endgültige Ziel der Ersatz des Tierversuchs ist, erscheint es uns als konzeptioneller Fehler wenn Erfolge in den anderen beiden Bereichen, welche das Potential haben, das Leiden einer riesigen Anzahl von Tieren zu verringern, außer Acht gelassen werden. Verbesserungs(=Refinement)ansätze verringern nicht notwendigerweise die Tierversuchszahlen, aber es wird diesen Tieren bedeutend weniger Leid zugefügt.

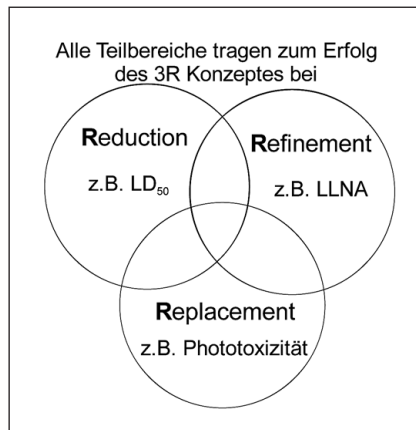


Abb. 5: Alle 3R Domänen vermindern Leiden und Schmerz

Zum Beispiel ist der Test für akute Toxizität (LD₅₀) ein „Reduction“ Test, der lokale Lymphknotentest (LLNA) ist eine akzeptierte „Refinement“ Methode und der *in vitro* Test mit Fibroblasten wird anerkannt, ohne dass *in vivo* Versuche nötig werden. Ein weiteres relevantes Beispiel für „Replacement“ wäre Hautkorrosion, wo *in vivo* Testrichtlinien durch *in vitro* Tests ersetzt wurden.

Da solche Nichtersatz(non-replacement)-Ansätze häufig außer Acht gelassen werden, möchten wir drei interessante Fälle aus diesem Bereich erwähnen. Das am häufigsten zitierte Beispiel für „Refinement“ ist der voll validierte und regulatorisch anerkannte (OECD Test Guideline 429, 2002) lokale Lymphknoten Test (LLNA) statt des Bühler Meerschweinchen Maximierungstests, um das Sensitivierungspotential von Chemikalien zu erfassen. Ein bekanntes Beispiel für „Reduction“ sind die neuen Tests zur Bestimmung von akuter Toxizität (OECD Test Guidelines TG420, 423, 425), welche seit dem 20. Dezember 2002 die alte LD₅₀ Richtlinie (TG401) ersetzen und die Anzahl Tiere um > 60-70% herabsetzen. Weniger bekannt sind die vielen Ansätze die als „intelligente Teststrategien“ bezeichnet werden und die ein riesiges Potential haben, den Tierversuch zu verringern. Ein anschauliches Beispiel ist die Teststrategie für akute Toxizität bei aquatischen Organismen (Daphnien, Algen, Fische), bei der durch die neue sequentielle Teststrategie, bei der die Fische zuletzt getestet werden, circa 60% der Fische eingespart werden (ESAC, 2006; Jeram, 2005). „Reduction“ Ansätze

sind unspektakulär hinsichtlich der entstehenden Publikationen (da sie ironischerweise den Grund zur Publikation beseitigen), sind aber hochgradig effektiv und, noch wichtiger, schnell in ihrer Implementierung und Auswirkung. Man kann ihren Wert in großen Programmen wie REACH kaum hoch genug einschätzen (z.B. mehr als eine Million gerettete Ratten durch verändertes Herangehen an die Bestimmung akuter Toxizität). In der groß angelegten Testung von Chemikalien (gemäß REACH), die in Europa bereits auf dem Markt sind, wird die Anzahl benötigter Tiere stark von der Anwendung von „read-across“ Ansätzen und dem intelligenten Umgang mit Information abhängen, und Millionen Tiere können durch intelligente Teststrategien gerettet werden. Das folgende Beispiel aus dem Bereich der Hautkorrosionstestung veranschaulicht die Stärke dieses Ansatzes: Die Entwicklung und Validierung der Ersatzmethode CORROSITEX (Corrositex, 2007) dauerte circa zehn Jahre. Diese Methode erkennt hauptsächlich Substanzen mit extremem pH. Daher wäre ein alternativer Ansatz eine Strategie, die automatisch (ohne *in vivo* oder *in vitro* Tests) Verbindungen mit extremem pH als korrosiv einstuft. Solch eine „intelligente“ Teststrategie (als Teil der geltenden OECD Richtlinien) verhindert extremes Leiden sofort, d.h. zehn Jahre Entwicklungszeit und zusätzliches Leiden werden eingespart.

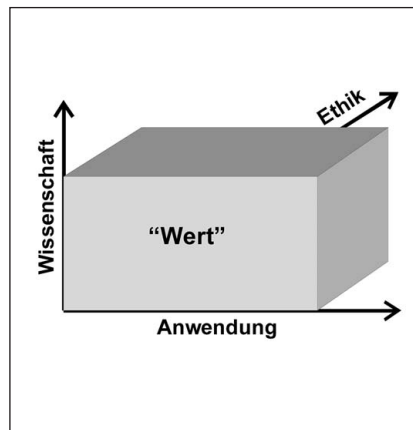


Abb. 6: Der Gesamtwert des 3R Ansatzes wird durch das Produkt verschiedener Dimensionen definiert

Denkfehler VI: Eindimensionale Betrachtung einzelner Wertedomänen

Wie bei jeder wissenschaftlichen Disziplin setzt sich der Gesamtwert der 3R Forschung aus verschiedenen Dimensionen, einschließlich „Qualität der Wissenschaft“, „Anwendbarkeit“ und „Ethik“, zusammen (Abb. 6). Im besonders multidisziplinären 3R Feld sind einige Gruppen stark auf eine der Dimensionen fokussiert und vernachlässigen häufig die anderen. Dies bewirkt, dass der Gesamterfolg der Disziplin unterschätzt wird und die 3R Forschung ihr volles Potential nicht immer erreicht. Die „wissenschaftliche Dimension“ ist auf die Relevanz und Kohärenz von Ersatzmethoden fokussiert, spielt aber auch eine wichtige Rolle bei



der Identifizierung und beim Herangehen an neue Herausforderungen und sollte immer wieder neue Arten von Lösungskonzepten sondieren. Nicht zu vernachlässigen in diesem Zusammenhang ist die inhärente Freude an guter Wissenschaft und die Faszination für dieses besonders interdisziplinäre Feld. Ein klares Bewusstsein des letzteren Punktes kann dazu beitragen mehr und mehr etablierte Forscher aus anderen Disziplinen sowie Studierende, die eine interessante Karriere suchen, in dieses Feld anzulocken. Die „Dimension der Anwendbarkeit“ beinhaltet Faktoren wie Kosten, Bedarf, Testqualität und Definition von Standards. Ein stärkeres Bewusstsein um den Wert dieser Domäne würde zu besserer Infrastruktur für solche Arbeiten führen, welche in vielen Ländern überhaupt nicht vorhanden und in den meisten anderen nur wenig entwickelt ist. Ein Ansatz, der in die richtige Richtung geht, ist die Etablierung eines Europäischen Referenzlabors (Community Reference Laboratory for Alternative Testing, CORRELATE, 2007) bei ECVAM. Doch auch hier erscheinen der Anwendungsbereich und vor allem die Finanzierung sehr gering im Vergleich zur riesigen zu bewältigenden Aufgabe und zum Wertpotential. Eine dritte Dimension beinhaltet die „ethischen Aspekte“ des Gebietes, wie zum Beispiel die Frage nach der Würde von Tieren, die Bilanzierung von Schmerz gegen mögliche Gewinne aus Tierexperimenten, die Definition des Wertes eines Tieres unabhängig von seinem Nutzen für den Menschen, das akzeptable Risiko für Menschen in Relation zu wirtschaftlichen Faktoren, den relativen Unterschied zwischen verschiedenen Tierespezies wie z.B. Primat versus Haustier versus Nager, usw. Großer zusätzlicher Wert und Fortschritt kann durch eine detaillierte Auseinandersetzung mit diesen Fragen anstelle eines dogmatischen und simplistischen Ansatzes gewonnen werden. (siehe folgendes Kapitel)

Denkfehler VII: Die Annahme dass die Anzahl der Versuchstiere mit deren Belastung korreliert

Dieser letzte hier besprochene konzeptionelle Fehler kann entweder zu einer Un-

ter- ODER Überschätzung des Wertes von 3R Methoden führen. Dieses Kapitel kann zusammen mit den folgenden Kapiteln eine Grundlage für weiterführende Diskussionen in diesem oder einem anderen Forum bilden.

In vielen Ländern, und auch in der EU, geben Tierversuchsstatistiken keine Information über das Leiden der Tiere wieder. Die Spanne von Schmerz und Leiden, die den Tieren bei Experimenten zugefügt werden können, ist extrem breit, und dies wird in keiner Weise berücksichtigt. Eine weitere Grauzone ist die Züchtung genetisch veränderter Tiere, was in den meisten Fällen nicht in der Tierversuchstatistik erfasst wird (im Gegensatz zur Generierung solcher Tiere). Ob die Züchtung genetisch veränderter Tiere für sie einen Stress darstellt findet bisher noch nicht genügend Beachtung, und die Einschätzung dieser Frage übersteigt manchmal die Kompetenzen derer, die die Tiere generiert, bzw. in Auftrag gegeben haben (Mertens and Rüllicke, 2007; Sauer et al., 2006).

Wenn man versucht, unter die Oberfläche statistischer Zusammenfassungen zu sehen erfährt man, wie schwer und oft unmöglich es ist, viel Information aus den Zahlen zu extrahieren. Zum Beispiel verbrauchte Griechenland 1999 keinen einzigen Fisch. Im Jahre 2002 wurden schon mehr als eine halbe Million verbraucht und diese Zahl hatte sich bis 2005 fast verdoppelt. Die Zahl ist jetzt höher als die Gesamtzahl der Fische, die insgesamt in den anderen 24 EU Mitgliedsstaaten verbraucht werden. Die Statistiken zeigen uns keinen Weg irgendeine Erklärung dafür zu finden. Als weiteres Beispiel möge eine scheinbar einfache Frage dienen: „Wie können wir die Situation von Haustieren (= companion animals) verbessern?“, d.h. wie können wir Methoden finden, die den Verbrauch von z.B. 24.000 Hunden pro Jahr im Tierversuch reduzieren können? Um dieses Problem anzugehen, müssen wir eine Antwort auf die Frage „Wo werden Hunde überhaupt als Versuchstiere eingesetzt?“ finden. Es gibt keine Möglichkeit, diese Frage mithilfe europäischer oder nationaler Statistiken zu beantworten. Wir können zwar darin finden, dass 500 Hunde für die Prüfung von Pestiziden verbraucht werden, 6.000 in biomedizinischer Forschung und ca.

13.000 für Sicherheitsprüfung von Medikamenten, aber es gibt keine weiteren Details. Diese Situation beeinflusst erheblich die Frage „Ist ein Test erfolgreich bei der Verbesserung der Situation von Hunden?“. Diese Frage ist schwer zu beantworten, wenn man nur offizielle Statistiken als Datengrundlage zur Verfügung hat. Die Beispiele zeigen, dass der Wert von Alternativmethoden teilweise schwer einzuschätzen ist, weil die Tierversuchstatistiken so schlecht nachvollziehbar und undurchsichtig sind.

Um ein weiteres Defizit der Statistiken zu verstehen wenden wir uns einer anderen Situation zu, nämlich dem „Draize Augenirritationstest“, der dringend eine Alternative erfordert (außer in Frankreich, wo der HET-CAM Test als Alternative voll akzeptiert ist). Eine Anzahl verschiedener Tests (z.B. isolierte Rinder-, Hühner-, oder Kaninchenaugen) können als Filtertests vorgeschaltet werden und, falls die Ergebnisse positiv sind, wird diese Information von den Behörden der EU und ihrer Mitgliedsstaaten akzeptiert. Laut den Statistiken der letzten 25 Jahre wurden ca. 7% aller getesteten Chemikalien als irritierend für das Auge eingestuft. Das würde bedeuten, dass alternative Methoden die Anzahl von Draize-Tests um ca. 7% reduzieren könnten. Ist dies beeindruckend oder eher vernachlässigbar? Wir brauchen darauf keine Antwort zu geben, denn die Frage selbst erscheint uns falsch formuliert. Es wäre besser zu wissen, wie viel Leiden verhindert wurde als statistische Information zur Anzahl durchgeführter Draize-Tests zu erfragen. Nach der Einführung von alternativen Tests wird immer noch eine hohe Anzahl von Augenirritationstests am Tierauge durchgeführt, aber nun größtenteils mit harmlosen Verbindungen. Starke Reizmittel werden durch die *in vitro* Tests meist schon herausgefiltert bevor sie in häufig sehr schmerzhaften Tierversuchen eingesetzt werden. Das veranschaulicht, dass eine große Erfolgsdomäne von Alternativmethoden momentan ihre Fähigkeit zur Vorselektion ist, wodurch die Verbindungen, die am giftigsten sind und Schmerzen hervorrufen, gar nicht erst in den Tierversuch kommen. Das wird allerdings auf keine Weise durch die Tierverbrauchsstatistiken wiedergegeben.

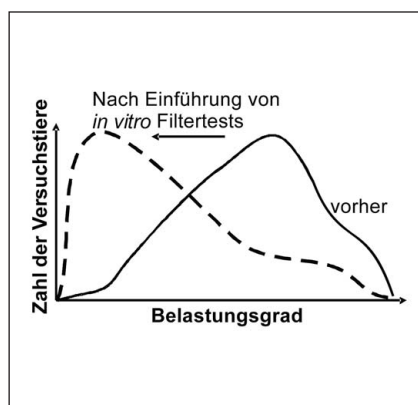


Abb. 7: Alternativmethoden als Filter können die Leidensverteilung von Tierversuchen verschieben

Die durchgezogene Linie zeigt eine hypothetische Leidensverteilung bei Tierversuchen vor der Anwendung von Alternativmethoden, um Verbindungen vorzusortieren bevor sie in Tieren eingesetzt werden. Die gestrichelte Linie zeigt die veränderte Verteilung nach Einsatz von Methoden, die als positive Filter anerkannt sind, d.h. wenn Verbindungen in diesen Tests positiv sind (die für Tiere schädlichsten Verbindungen), werden sie nicht an Tieren getestet.

Zuletzt wollen wir noch eine Angelegenheit ansprechen, die häufig außer Acht gelassen wird, aber nicht übersehen werden sollte: Die Unterschätzung der Wirkung von Alternativmethoden auf den Forschungsdurchsatz. Einige Gebiete der Medikamentenentwicklung sind in ihren Dimensionen noch eingeschränkt, weil die Tierversuche einen Engpass darstellen. Ein gutes *in vitro* System kann die Anzahl Firmen, die an einem Gebiet Interesse haben, vergrößern wie auch die Anzahl von Verbindungen, die in jeder Firma getestet werden. Schlussendlich kann das auch eine viel größere Anzahl von Tieren in pharmakologischen, Sicherheits- und Qualitätsprüfungen erfordern als je zuvor. Aber parallel dazu werden weitere Alternativmethoden entwickelt. Daher wissen wir momentan nicht, wie der ethische Wert solcher Alternativmethoden eingeschätzt werden sollte.

Wie könnte der Fortschritt im Bereich der 3R Methoden sonst bewertet werden?

Wenn uns die heutigen Statistiken nicht weiterhelfen, wie sollen wir dann eine sinnvolle Einschätzung des Wertes von 3R Methoden erlangen? Es scheint uns wichtig, eine Alternative zur negativen Definition anhand von Tierversuchstatistiken zu finden und dem Feld lieber eine positive Basis zu geben. Man könnte daher fragen: „Wie groß ist der Output von Alternativmethoden?“, „Wie viel haben 3R Methoden und 3R Ansätze zum Gesamtoutput (Forschung, Sicherheitsprüfung, Krankheitsmechanismen) beigetragen?“ oder „Wäre das gleiche Endergebnis auch ohne *in vitro* Methoden erreicht worden, und wenn ja, gleich gut?“ Man muss auch die steigende Anzahl von Menschen die auf diesem Gebiet arbeiten berücksichtigen, die hier Karrieremöglichkeiten und wissenschaftliche Herausforderungen finden. Dabei ist es notwendig, alle Anwendungsdomänen zu berücksichtigen und die gesamte Breite des Feldes über die Toxikologie hinaus widerzuspiegeln, wie es zum Beispiel die schweizerische Forschung 3R oder die Doerenkamp-Zbinden Stiftung tun (Mair und Wick, 2007; DZF, 2007).

Man könnte 3R auch auf medizinischer oder wissenschaftlicher Basis beurteilen und fragen: „Wie viel besser sind *in vitro* Modelle in der Vorhersage von Sicherheit oder pharmakologischer Wirkung im Menschen geworden?“ Gleichsam könnte man erfragen, wie viel besser unser mechanistisches Verständnis wichtiger Prozesse in der Toxikologie und medizinischen Forschung geworden ist. Obwohl dieser Aspekt besonders wichtig erscheint, darf man dabei nicht vergessen, dass nicht alle *in vitro* Forschung auch 3R Forschung ist und umgekehrt. Ein weiterer Wert des 3R Ansatzes, der nicht außer Acht gelassen werden sollte, besteht in der Zeit und dem Geld, die durch Alternativmethoden eingespart werden. Wir haben auch in vielen anderen Bereichen schon gesehen, dass Wirtschaftlichkeit nicht völlig von ethischen, ökologischen und moralischen Überlegungen entkoppelt werden kann, sondern sogar eine wertvolle Triebkraft sein kann. Andererseits ist die ethische Dimension selbst

nicht nur Triebkraft des Feldes, sondern eignet sich auch als alternative Wertbasis: Neue Methoden der Bewertung sollten unbedingt das Ausmaß, in dem Stress und Leiden vermindert wurden, beinhalten, nicht nur als statistische Durchschnittswerte, sondern auch als viele individuelle Fallbeispiele.

Wege aus dem Dilemma

Eine breitere Bewertungsbasis ist eine ideale Plattform für eine große Vielfalt an Ansätzen, durch die die 3R Forschung ihre Außen- aber auch Innenwahrnehmung verbessern kann. Die Veränderung dieser Wahrnehmung ist nicht nur von kosmetischem Wert, sondern ist wichtig, um das Gebiet konzeptionell und technisch vorwärts zu bringen. Solche Fortschritte werden wiederum Auswirkungen in angrenzenden und überlappenden Feldern wie der Toxikologie, Stammzellforschung, Pharmakologie und medizinischen Forschung haben. Ansätze in diese Richtung können in Aktionspakete zusammengefasst werden.

Aktionspaket eins würde zum Beispiel Ansätze zur Verbesserung von Tierversuchstatistiken und ihrer Verwendung beinhalten. Statistiken und Datenbanken sollten transparenter und besser nachzuvollziehen sein, freien Zugriff gewähren und anwenderfreundlich sein, und sie sollten auf jeden Fall Maße für Stress und Leiden beinhalten - zusammen mit Information zum Hintergrund der Tierversuche. Ansätze die allein auf Tierzahlen beruhen, wie sie oft beim Vergleich verschiedener REACH Szenarien verwendet werden, erscheinen zynisch und lenken von den eigentlichen Problemen nur ab.

Aktionspaket zwei könnte sich mit der unausgewogenen Publikationslage und dem Mangel an wichtiger Information aus der Industrie beschäftigen. Anreize für Fallstudien und Publikationen könnten aus der Industrie gegeben werden. Ein gutes Beispiel für den Nutzen solcher Publikationen stellt eine Studie in der Firma Organon zum Effekt von 3R Methoden innerhalb der eigenen Forschung dar (Verbost et al., 2007). „Veröffentlichung“ wird hier im breiteren Sinne definiert, einschließlich der Einspeisung von Daten in frei



zugängliche Datenbanken, der Unterstützung von *in vivo-in vitro* Vergleichen und der Herausgabe von Information zur Existenz und der praktischen Anwendung von Alternativmethoden. Dieser Prozess kann durch die Einrichtung größerer nationaler oder internationaler Zentren, die dem Aufbau von Infrastruktur für die Entwicklung von Alternativmethoden gewidmet sind, entscheidend vorangetrieben werden. Große Organisationen wie EPAA haben bereits mit der Arbeit an diesem Aktionspaket begonnen.

Aktionspaket drei würde durch die Fokussierung auf eine Anzahl Zentren, die für Referenzsubstanzen und entsprechende Datenbanken verantwortlich wären, in diese Richtung weiter gehen. Solche Schritte mögen trivial erscheinen. Ihre Notwendigkeit zeigt jedoch wie unterentwickelt die Infrastruktur in diesem Feld noch ist und wo dringend etwas getan werden muss. Stellen Sie sich vor, sie suchen eine Liste von Verbindungen, um ein Modell, z.B. für Entwicklungsneurotoxizität, zu validieren. Sie wundern sich, warum Sie keine solche Liste in der Literatur finden können. Die Erklärung ist ganz einfach. Gute, validierte Listen solcher Verbindungen sind noch nicht verfügbar. Auch wenn Sie eine Liste der relevanten Verbindungen kennen würden, wäre es für Sie äußerst schwierig, manchmal unmöglich, diese auch zu bekommen. Viel könnte hier mit relativ wenig Einsatz erreicht werden, und einige dieser Anläufe wurden bereits initiiert, zum Beispiel von ECVAM.

Aktionspaket vier beinhaltet ganz andersartige Aktivitäten, und fokussiert sich auf die Einigkeit und die Grundstärken des 3R Gebiets. Mit dem Wachstum des Gebiets wird es wichtig sein, zwischen den verschiedenen Wertedomänen und „R“-Domänen guten Kontakt zu halten und Interaktionen zwischen ihnen zu fördern; etwas, was zu den wichtigsten Verdiensten, z.B. des Linzer Kongresses zu Alternativen zählt, der vom Zentrum

für Ersatzmethoden zum Tierversuch (ZET) organisiert wird. Eine große Herausforderung ist auch die Notwendige Öffnung der Grenzen zu anderen Disziplinen. Einerseits wird hier eine gewisse Selbstkontrolle notwendig sein, um zwischen Forschung an Alternativmethoden und reiner mechanistischer oder *in vitro* Forschung zu trennen. Andererseits muss man dabei sehr vorsichtig sein, nicht den Fehler zu begehen, das Gebiet zu eng zu definieren und dadurch Nachbardisziplinen auszuschließen. Solch eine Herangehensweise trüge das Risiko, dass das Gebiet verarmt, wie schon vorher in der Toxikologie geschehen (Lotti and Nicotera, 2002).

Aktionspaket fünf beinhaltet alle Ansätze, um einen fairen Vergleich zwischen Alternativmethoden und klassischen Tierexperimenten anzustellen. Vor allem im Bereich Toxikologie bedeutet dies vergleichbare Validierungskriterien sowohl für die Tierexperimente als auch für die Alternativmethoden anzuwenden und das Abwenden vom Dogma, dass der Tierversuch immer der Goldstandard ist. Ebenso sollte man verlangen, dass mehr Wert auf Relevanz-basierte und mechanistisch-charakterisierte Tierversuche gelegt wird (im Gegensatz zu rein historisch und durch Behördenforderungen legitimierten Versuchen), welche unter stringenteren Kriterien validiert worden sind, wie sie z.B. in der Evidenz-basierten Medizin angewendet werden. Eine neue Bewegung in diese Richtung nennt sich „Evidenz-basierte Toxikologie“ (EBT) (Guzelian et al. 2005, Hoffmann and Hartung, 2006). Diese Denkweise wäre ein Weg in die Zukunft, der Alternativmethoden ihren angemessenen Platz und Wert zuteil werden ließe.

Aktionspaket sechs führt diesen Gedankengang konsequent zum Abschluss und beendet die Vorherrschaft des „Tierversuchsspiels“ endgültig. In einem Grundsatzdokument „Toxicity testing in the 21st century“ (Toxizitätstests im

21sten Jahrhundert) definierte der National Research Council der USA (NRC, 2007) eine Vision, in der das „S3R Spiel“ eine Hauptrolle besetzt. Die Regelsätze des Tierversuchsspiels und des 3R Spiels werden hier bezüglich ihrer Vorherrschaft vertauscht. Statt klassischerweise Tiere als „Black Box“ zu sehen und dann nachzuforschen was in den Fällen einer auftretenden Toxizität eigentlich passiert, wird ein Bottom-up-Ansatz zum mechanistischen Verständnis gefordert. Mechanistische Forschung mit Alternativmethoden kann die zentralen Abläufe, die vielen Formen von Toxizität gemein sind, definieren. Toxische Verbindungen würden dann zunächst durch die Prozesse die sie auslösen charakterisiert werden, und Tierversuche würden erst in späteren Stadien der Integration durchgeführt werden, um diese Informationen zu ergänzen und Lücken zu schließen. Die starke Interdisziplinarität und die Einbettung des 3R Gebietes in die biomedizinische Forschung sind eine ideale Grundlage für diesen Ansatz (Lotti and Nicotera, 2002). In dieser neuen Form der Mechanismus-basierten Toxikologie kann 3R Forschung ihren vollen Wert für Tiere, für die Wissenschaft und für die Menschheit erreichen.

Literatur

Literaturverzeichnis siehe „References“ Seite 23f

Korrespondenzadresse

Prof. Dr. Marcel Leist
Doerenkamp-Zbinden Lehrstuhl
für *in vitro* Methoden zum
Tierversuchersatz
Universität Konstanz
Postfach M657
78457 Konstanz
Deutschland
Tel.: +49-7531-885037
Fax: +49-7531-885039
E-mail: marcel.leist@uni-konstanz.de