

Die Genehmigung von Tierversuchen nach Richtlinie 2010/63/EU – zentrale Konzepte zur Prüfung der „ethischen Vertretbarkeit“, ihre Funktion, praktische Schwierigkeiten und die Beschränkung ihrer „ethischen Reichweite“

Matthias Eggel

Weltweit werden Jahr für Jahr ca. 115 Millionen Tiere (Andrew Knight 2008) zu wissenschaftlichen Zwecken im Namen der Forschung verwendet mit dem Ziel, wichtige Erkenntnisse für die Gesundheit von Mensch und Tier und für den Schutz der Umwelt zu erzielen.

Diese Praxis wirft eine Vielzahl von ethischen Fragen auf, polarisiert die Gesellschaft und wird auch akademisch kontrovers diskutiert. Tiere sind in unserer moralischen Gesellschaft angekommen und die Zusprechung von moralischem Status wirft Fragen zur Mensch-Tier-Beziehung generell und zu den Tierversuchen im Speziellen auf. Welche Handlungen an Tieren sind zu wissenschaftlichen Zwecken erlaubt, geboten oder verboten? Ist es ethisch vertretbar, dass Tiere Belastungen erfahren, damit der Mensch Erkenntnisse zur Verbesserung der menschlichen Gesundheit erzielt? Wie und unter welchen Umständen kann tierliches Leid und die „Opferung“ tierlicher Interessen für menschliche Interessen „ethisch“ gerechtfertigt werden? Diese Fragen sind nicht neu, aber immer noch (und vielleicht mehr denn je) Bestandteil des akademischen Diskurses (z. B. Belastungsobergrenze, Primaten in der Forschung, Reproduktionskrise) und Teil einer andauernden öffentlichen Debatte (z. B. Initiativen zum Verbot von Tierversuchen).

Bedenken im Zusammenhang mit Tierversuchen können auch auf gesetzlicher Ebene beobachtet werden. Tierversuche sind in den EU-Staaten durch die Richtlinie 2010/63/EU „zum Schutz der für wissenschaftliche Zwecke verwendeten Tiere“ des Europäischen Parlaments und des Rates der Europäischen Union streng reglementiert und bewilligungspflichtig. Die Verwendung von Tieren zu Forschungszwecken wird auf Bereiche eingeschränkt, die letztlich einen „Nutzen für die Gesundheit von Mensch, Tier und Umwelt nach sich ziehen können“ (Erwägungsgrund 12).

Im Folgenden versucht dieser Aufsatz, eine (grobe) Übersicht über das Genehmigungsverfahren und die Prüfung auf „ethische Vertretbarkeit“ nach Richtlinie 2010/63/EU zu geben. Diesbezüglich wird die Prüfung auf

Eignung, wissenschaftliche Erforderlichkeit und Verhältnismäßigkeit erörtert. Durch die Diskussion der Konzepte, auf welchen diese Begriffe beruhen, und der Funktionen, die sie innerhalb der Genehmigung von Tierversuchen einnehmen, soll ein Teil der normativen Infrastruktur der Richtlinie nachgezeichnet werden. Punktuell werden (in der gebotenen Kürze und ohne Anspruch auf Vollständigkeit) praktische Schwierigkeiten und Grenzen (der „ethischen Reichweite“) skizziert, die sich aus diesen Konzepten und ihren Funktionen ergeben.

1. *Grundlegende Aspekte der gesetzlichen Regelung von Tierversuchen nach Richtlinie 2010/63/EU*

Laut Richtlinie 2010/63/EU zum Schutz der für wissenschaftliche Zwecke verwendeten Tiere ist unter Tierversuch (oder „Verfahren“)

„jede invasive oder nicht invasive Verwendung eines Tieres zu Versuchszwecken oder anderen wissenschaftlichen Zwecken [...], die bei dem Tier Schmerzen, Leiden, Ängste oder dauerhafte Schäden in einem Ausmaß verursachen kann, das dem eines Kanüleneinstichs gemäß guter tierärztlicher Praxis gleichkommt oder darüber hinausgeht“ (Art. 3),

zu verstehen. Die Richtlinie spricht empfindungsfähigen Tieren einen „intrinsischen Wert“ zu,

„der respektiert werden muss und aus welchem Grund Tiere stets als fühlende Wesen behandelt werden, und ihre Verwendung in Verfahren auf Bereiche beschränkt werden sollte, die letztendlich einen Nutzen für die Gesundheit von Mensch und Tier oder für die Umwelt nach sich ziehen können“ (Erwägungsgrund 12).

Diesbezüglich inkludiert die Richtlinie lebende nichtmenschliche Wirbeltiere (einschließlich selbständig Nahrung aufnehmende Larven und Föten von Säugetieren ab dem letzten Drittel ihrer normalen Entwicklung) und lebende Kopffüßer (Art. 1, Abs. 3). Insekten (z. B. *Drosophila Melanogaster*), Würmer (z. B. *Caenorhabditis Elegans*) und andere nicht-menschliche Tiere, denen die Empfindungsfähigkeit abgesprochen wird, werden nicht berücksichtigt.

In Art. 5 schränkt die Richtlinie die Verwendung von Tieren zu Forschungszwecken ausschließlich auf die folgenden Zwecke ein: Grundlagenforschung, translationale oder angewandte Forschung, regulatorische Versuche, Schutz der Umwelt im Interesse der Gesundheit und Wohlerge-

hen von Mensch und Tier, Forschung zur Erhaltung der Arten, Ausbildung und forensische Untersuchungen.

Für eine Genehmigung eines Tierversuchs muss ein Antrag gestellt werden, der von den zuständigen Behörden auf verschiedene Kriterien geprüft wird (Art. 37-38). Dabei ist die Prüfung von Eignung, wissenschaftlicher Erforderlichkeit und Verhältnismäßigkeit zentral.

2. Prüfung der Eignung, wissenschaftlicher Erforderlichkeit und Verhältnismäßigkeit

2.1 Prüfung der Eignung

Artikel 38/a der Richtlinie verlangt „eine Beurteilung der Projektziele, des erwarteten wissenschaftlichen Nutzens oder des pädagogischen Werts“. Diese Beurteilung wird auch als Prüfung der Eignung bezeichnet (Swiss Academies of Arts and Sciences 2017). Es sollen die Fragestellung und das Ziel auf Plausibilität und die beschriebenen Methoden auf ihre Durchführbarkeit und Erfolgsaussichten geprüft werden. Ist die Hypothese durch genügend Literatur gestützt? Ist die Wahl des Modells geeignet, um die Fragestellung zu beantworten? Wie wird die Fragestellung messbar gemacht? Ist das experimentelle Design detailliert genug beschrieben bezüglich der Kriterien der guten wissenschaftlichen Praxis, d. h. Verblindung, Randomisierung, Stichprobengröße und statistische Auswertung? Wie steht es um die Reproduzierbarkeit und Generalisierbarkeit der erwarteten Ergebnisse etc.? Ziel dieser Prüfung ist es sicherzustellen, dass der vorgeschlagene Tierversuch das geeignete „Instrument“ ist, um ein gegebenes Versuchsziel zu erreichen.

Die hier zugrundeliegende Annahme ist, dass nur Forschung, die höchsten wissenschaftlichen Standards gerecht wird, belastbare, reproduzierbare und generalisierbare Daten generieren kann und dass die Verwendung von Tieren zu Versuchszwecken nur unter dieser Bedingung ethisch vertretbar ist.

Praktische Probleme in der Prüfung der Eignung

Bezüglich der Qualität der Forschung sind in den letzten Jahren im Zuge der sogenannten „Reproduzierbarkeitskrise“ Zweifel aufgekommen. Mangelnde Reproduzierbarkeit von Forschungsergebnissen wird mit als Grund

genannt für die geringe Übertragbarkeit der Forschungsergebnisse aus Tierversuchen in die klinische Praxis.¹ Verschiedene Studien zeigen auf, dass ein signifikanter Anteil von veröffentlichten Forschungsergebnissen aufgrund mangelnder Qualität der Forschung und ihrer Veröffentlichungspraktiken (Ismail Kola/John Landis 2004; Emily S. Sena et al. 2010; H. Bart van der Worp et al. 2010) unzuverlässig sind. Zudem enthalten viele der untersuchten Studien nur ungenügend oder gar keine Informationen zu Maßnahmen zur Vermeidung von systematischen Fehlern (d. h. Verblindung, Randomisierung, Berechnung der Stichprobengröße, geeignete Kontrollgruppen, Plan für statistische Auswertung etc.). Ein Teil dieses Problems wurde auch in der Prüfung der Eignung im Genehmigungsverfahren verortet. Eine Untersuchung Schweizer Tierversuchsanträge kam zum Schluss, dass die Behörden Anträge oft auf Basis von Vertrauen anstelle von Beweisen für die wissenschaftliche Qualität von Forschungsprojekten genehmigen (Lucile Vogt et al. 2016). Dies passiert nicht zuletzt deshalb, weil den Behörden oft wichtige Informationen über den experimentellen Aufbau und über Maßnahmen zur Vermeidung von systematischen Fehlern fehlen. Aktuell beinhalten Genehmigungsverfahren keine oder nur eine ungenügende systematische Prüfung der wissenschaftlichen Qualität. Im Lichte der Reproduzierbarkeitskrise und der mangelhaften Forschungsqualität scheint es aber dringend notwendig, dass Methoden zur systematischen Prüfung der wissenschaftlichen Qualität in die Genehmigungspraxis integriert werden. Forscher und Behörden sind sich dieser Mängel bewusst und sehen Handlungsbedarf. Diesbezüglich hat z. B. Hanno Würbel (2017) vorgeschlagen, Projektanträge systematisch auf die 3V, d. h. die drei Validitäten (*interne*, *externe* und *Konstruktvalidität*) zu prüfen.² Würbels Vorschlag wurde 2020 von den Schweizer Behörden im Zu-

-
- 1 Eine Studie zeigt, dass von 500 neuroprotektiven Behandlungen, welche im Tiermodell effektiv waren, keine einzige auch im Menschen eine Wirkung gezeigt hat (Victoria E. O'Collins et al. 2006). Eine andere Studie zu Schlaganfall konnte zeigen, dass von 100 Behandlungsmethoden, welche im Tierversuch einen Effekt gezeigt hatten, nur eine einzige auch im Menschen wirksam war (ebd.).
 - 2 Validität stellt ein Kriterium für die Überprüfung der Qualität und Güte von Modellen, Mess- und Testverfahren dar. *Interne Validität* beschreibt den Grad der Kausalkette zwischen Experiment und Ergebnis, d. h., inwieweit die Planung, Durchführung und Auswertung des Experiments die Möglichkeit systematischer Fehler ausschließt. Dies beinhaltet z. B. die Inklusion von Verblindung und Randomisierung, eine adäquate Stichprobenberechnung, geeignete Kontrollgruppen und einen geeigneten Plan für die statistische Auswertung (Hanno Würbel 2017). Die *externe Validität* besagt, inwieweit die Ergebnisse eines Tierversuchs auf andere Populationen von Tieren (und/oder einschließlich Menschen) und/oder andere Um-

ge der Überarbeitung des Antragsformulars für Tierversuche übernommen. Somit beinhaltet das Schweizer Prüfungsverfahren (im Gegensatz zu den meisten europäischen Prüfungsverfahren) seither eine systematische Evaluierung der 3V. Analog haben Daniel Strech und Ulrich Dirnagl (2019) vorgeschlagen, Projektanträge auf *Robustness*, *Reporting* und *Registration* zu untersuchen. *Robustness* bezieht sich auf eine systematische Prüfung der wissenschaftlichen Qualität (wie sie Würbel vorgeschlagen hat). Oft fehlen in Publikationen wichtige Informationen zu Versuchsaufbau, Durchführung und Analyse. Diesbezüglich verlangt *Reporting*, die transparente und detaillierte Publikation von Forschungsergebnissen, als weiteren unerlässlichen Faktor die Reproduzierbarkeit zu verbessern. Nur so können nämlich Studien auf ihre wissenschaftliche Qualität und methodische Zuverlässigkeit geprüft werden, und nur so ist es anderen Forschern möglich, Experimente zu wiederholen und auf publizierten Arbeiten aufzubauen. *Registration* verlangt das Hochladen eines detaillierten Forschungsplans für einen geplanten Tierversuch (ähnlich klinischen Versuchen) auf Online-Plattformen vor seiner Durchführung. Dies soll verhindern, dass Resultate gar nicht oder nur selektiv publiziert werden (*publication bias*).³ Durch die Implementierung dieser vorgeschlagenen Methoden soll zum einen die wissenschaftliche Qualität, zum anderen aber auch die Qualität der Prüfung auf Eignung verbessert werden.

weltbedingungen übertragbar sind. Dies beinhaltet das Aufteilen von Experimenten in mehrere unabhängige Replikate (Richard Paylor 2009; Natasha A. Karp et al. 2014), systematische Heterogenisierung relevanter Variablen (z. B. Tierarten/ Tierstämme, Haltungsbedingungen [S. Helene Richter et al. 2010 und 2011]) oder Implementierung von sogenannten *multilaboratory studies* (Rachel Wodarski et al. 2016). Niedrige externe Validität kann ein Grund sein für ungenügende Reproduzierbarkeit und Translatierbarkeit von Resultaten. Die *Konstruktvalidität* (Hanno Würbel 2017) gibt an, wie gut ein Test das misst, was er zu messen vorgibt. Dies beinhaltet z. B. die Wahl eines geeigneten Tiermodells, um eine spezifische Fragestellung zu beantworten, z. B. Grad der Übereinstimmung zwischen einem Tiermodell für eine spezifische Krebsart und der zu messenden Pathophysiologie – d. h. in diesem Falle die zu untersuchende Art von Krebs – im Menschen.

- 3 An dieser Stelle sind auch die PREPARE- und ARRIVE-Richtlinien zur Verbesserung von Planung, Durchführung und Dissemination von Tierversuchen und deren Ergebnisse zu erwähnen (PREPARE = *Planning Research and Experimental Procedures on Animals: Recommendations for Excellence* [Adrian Smith et al. 2018]; ARRIVE = *Animal Research: Reporting of In Vivo Experiments* [Carol Kilkenny et al. 2010]). Wissen um diese und Gebrauch dieser Richtlinien scheinen aber noch relativ gering zu sein, hat aber in den letzten Jahren zumindest zugenommen (David Baker et al. 2014; Thomas S. Reichlin et al. 2016; Kaitlyn Hair et al. 2019).

Laut Gesetz ist nur Forschung ethisch vertretbar, die höchsten wissenschaftlichen Standards genügt, weil nur solche Forschung reproduzierbare und generalisierbare Daten generieren kann. Dies ist auch im Zusammenhang mit biomedizinischer Forschung von zentraler Bedeutung, denn nur reproduzierbare und generalisierbare Daten haben das Potential, auf den Menschen übertragbar zu sein, und sind dementsprechend eine notwendige Bedingung für die Realisierung von praktischem Nutzen (d. h. für die Förderung menschlicher Gesundheit). Als solches ist die Prüfung auf Eignung der erste Prüfschritt (und die erste notwendige Bedingung) für die ethische Vertretbarkeit von Tierversuchen. Mit Bezug auf die Reproduzierbarkeitskrise und die angesprochenen Mängel im Genehmigungsverfahren wird die momentane Praxis dieser Aufgabe in vielen europäischen Ländern noch unzureichend gerecht und muss verbessert werden. Zudem ist eine Prüfung der Eignung eines Tierversuchs auch eine notwendige Bedingung für die Prüfung der wissenschaftlichen Erforderlichkeit (als zweiter Prüfschritt und zweite notwendige Bedingung für die ethische Vertretbarkeit von Tierversuchen). Diese soll im nächsten Abschnitt erläutert werden.

2.2 Prüfung der wissenschaftlichen Erforderlichkeit

Nächster zentraler Bestandteil der Genehmigung von Tierversuchen ist die Prüfung der wissenschaftlichen Erforderlichkeit. Ziel dieser Prüfung ist es sicherzustellen, dass der geplante Versuch und dessen Ziel zwingend und ausschließlich durch einen Tierversuch durchgeführt und erreicht werden können. Dieser Prüfung liegt die Annahme zugrunde, dass ausschließlich Tierversuche ethisch vertretbar sind, die nicht durch Alternativmethoden ersetzt werden können, und dass die Belastungen an Versuchstieren so gering wie möglich zu halten sind. Diese Idee geht auf William Russell und Rex Burch und ihr einflussreiches Buch *The Principles of Humane Experimental Technique* zurück, in welchem sie 1959 das 3R-Prinzip beschrieben haben. Die 3R stehen für *Replacement*, *Reduction* und *Refinement*.⁴ *Replacement* verlangt, Tierversuche, wann immer möglich, zu vermeiden und auf Alternativmethoden auszuweichen, für welche keine Tiere verwendet wer-

4 Auf Deutsch auch als „Vermeidung, Verminderung und Verbesserung“ bekannt. Wegen der Verwechslungsgefahr mit dem 3V-Prinzip, welches von Hanno Würbel (2017) beschrieben wurde, werden hier die englischen Anfangsbuchstaben der 3R verwendet.

den müssen, d. h. Computersimulationen, mathematische Modelle, Stammzellen (z. B. Organoide), Zellkulturen, *Organ-on-a-chip* etc. Falls der Erkenntnisgewinn durch solche Alternativmethoden gewonnen werden kann, ist die Verwendung von Versuchstieren nicht notwendig und deshalb verboten. Die Richtlinie, die in Artikel 38/b eine „Bewertung des Projekts im Hinblick auf die Erfüllung der Anforderung der Vermeidung, Verminderung und Verbesserung“ vorschreibt, bezeichnet es (im Erwägungsgrund 10) dementsprechend auch als finales (langfristiges) Ziel, ganz auf die Verwendung von Tieren in der Forschung verzichten zu können. *Reduction* bezeichnet Bemühungen, die Anzahl der verwendeten Tiere auf das Mindestmaß zu reduzieren, welches immer noch valide Daten ermöglicht (z. B. genügend statistische Robustheit und Wahrscheinlichkeit, Behandlungseffekte detektieren zu können). Diesem Prinzip liegt der Gedanke zugrunde, dass sowohl die Verwendung von zu vielen, aber auch von zu wenigen Versuchstieren ethisch problematisch ist. Könnte man den Erkenntnisgewinn mit weniger Versuchstieren erreichen, würde die Verwendung der überschüssigen Tiere (und das überschüssige Leiden dieser Tiere) keinen zusätzlichen Nutzen bringen und wäre somit ethisch unzulässig. Werden hingegen zu wenige Tiere verwendet, kann die statistische Auswertung und damit die Aussagekraft eines Versuchs so sehr eingeschränkt sein, dass keine validen Daten mehr gewonnen werden können und dementsprechend Tiere unnützer- und unnötigerweise geopfert wurden. *Refinement* verlangt die Förderung des Wohlbefindens der Tiere sowie die Ausführung von Verfahren mit der kleinstmöglichen Belastung für die Versuchstiere. Dementsprechend sind Belastungen für Versuchstiere nur zulässig, wenn sie für den Versuch wissenschaftlich absolut notwendig sind.⁵

Praktische Probleme der Prüfung der Erforderlichkeit

Die 3R mit ihrer mittlerweile 60-jährigen Geschichte sind weithin als „ethisches Leitkonzept“ zur Leidverminderung von Versuchstieren akzeptiert. Die Implementierung der 3R bereitet aber in der Praxis weiterhin Schwierigkeiten. Im Zusammenhang mit *Reduction* z. B. mit dem Ziel, nicht mehr Tiere als notwendig zu verwenden, ist es in der Wissenschaft

5 Ein solcher Standpunkt, wie ihn die Richtlinie mit ihrem Fokus auf die 3R vertritt, welcher Tierversuche an sich ethisch vertretbar findet, aber für die Verbesserung und „Optimierung“ der Haltungsbedingungen und für die Leidvermeidung einsteht, wird oft als *animal welfare*-Position benannt (Gary L. Francione 1996).

zum Standard geworden, die genetischen Versuchsbedingungen sowie die Umwelt-Faktoren rigoros zu homogenisieren und standardisieren. Diese rigorose Standardisierung ermöglicht es zwar, die Test-Sensitivität zu erhöhen und somit die Zahl der für einen Versuch benötigten Tiere zu verkleinern (S. Helene Richter et al. 2009). Dies ist aber problematisch, weil eine (zu) rigorose Standardisierung zwar die Aussagekraft für die untersuchte Population und für die gegebenen Versuchsbedingungen (i. e. hohe interne Validität) steigern kann, aber gleichzeitig die Übertragbarkeit über die Stichprobe und Versuchsbedingungen hinaus (z. B. externe Validität sinkt) reduziert wird. Dieses Problem wurde als *Standardization fallacy* bekannt (S. Helene Richter et al. 2009; Hanno Würbel 2000). Diesbezüglich wurden Methoden zur kontrollierten Heterogenisierung von Versuchsbedingungen entwickelt, um der zu rigorosen Standardisierung entgegen zu wirken und die Reproduzierbarkeit und Generalisierbarkeit von Forschungsergebnissen zu verbessern.⁶ Unabhängig von Standardisierung und Heterogenisierung von Versuchsbedingungen bergen zu kleine Stichproben (teilweise motiviert durch den *Reduction*-Gedanken) und die daraus folgende ungenügende statistische Aussagekraft die (wissenschaftlich und ethisch relevante) Gefahr, dass Behandlungseffekte nicht gemessen oder gemessene Behandlungseffekte überschätzt werden und dementsprechend die Reproduzierbarkeit von Resultaten einschränken können (Katherine S. Button et al. 2013).⁷ Wie diese Beispiele aufzeigen, kann der *Reduction*-Gedanke, wenn richtig angewandt, vermeidliches Tierleid minimieren, aber bei der (zu rigorosen) Minimierung der verwendeten Tierzahl ist Vorsicht geboten.

Bezüglich *Refinement* haben diverse Studien den Einfluss von Haltungsbedingungen, der Handhabung (*handling*) von Versuchstieren sowie der Durchführung eines Tierversuchs (z. B. Operationen) auf das Wohlergehen der Versuchstiere aufgezeigt. Und obwohl das Ausschöpfen der 3R gesetzlich vorgegeben ist und das Wohlbefinden von Versuchstieren aus ethischer wie auch wissenschaftlicher Perspektive wichtig ist⁸, sehen verschiedene Publikationen Handlungsbedarf bei der konsequenten Implementierung von *Refinement*. So zeigen z. B. Studien, dass Standardbehausungen die Entwicklung und Funktion des Gehirns in Mäusen negativ beeinflussen können (Hanno Würbel 2001), dass Käfige nicht immer groß genug

6 Siehe oben, Fußnote 2.

7 Katherine S. Button et al. (2013) haben diesbezüglich gezeigt, dass das Problem der statistischen Aussagekraft in den Neurowissenschaften erstaunlich tief ist.

8 Unnötige Belastungen können Resultate verfälschen und somit die Validität einer Studie gefährden.

sind, damit Ratten stehen können (was wichtig für ihr Wohlbefinden ist) (Hibba Mazhary/Penny Hawkins 2019), oder dass Behausungen zu wenig „angereichert“ sind, um arten-spezifisches Verhalten genügend zu ermöglichen (z. B. Erkunden, Spielen, Nestbau etc.). Zudem wurde gezeigt, dass neueste wissenschaftliche Erkenntnisse für das schonende *handling* der Tiere, z. B. das Ersetzen des *tail handlings* (das Ergreifen der Maus am Schwanz) durch das Aufgreifen mit Hilfe eines Tunnels, noch ungenügend umgesetzt werden (Lindsay J. Henderson et al. 2020) oder dass die Durchführung von Operationen⁹, die post-operative Schmerzbehandlung (Kathrin Herrmann/Paul Flecknell 2018 a) sowie die Wahl der Tötungsmethoden laut verschiedenen Berichten noch verbessert werden können (Kathrin Herrmann/Paul Flecknell 2018 b). Die Inklusion neuer wissenschaftlicher Erkenntnisse und Methoden in die Praxis (von Forschung, Training und Lehre, aber auch von Prüfungsverfahren) erfolgt in der Regel mit einer gewissen zeitlichen Verzögerung. Nichtsdestotrotz werfen die genannten Beispiele die Frage auf, ob die 3R und die Verbesserung gängiger Praktiken tatsächlich immer ausgeschöpft werden oder ob in gewissen Fällen die Aufnahme und Implementierung neuer Methoden (und die damit verbundenen ethischen Bedenken der 3R zum Schutz von Tieren und deren Leidverminderung) durch praktische oder ökonomische Interessen übertrumpft werden. Ist dies tatsächlich der Fall, dann leiden Tiere unnötigerweise, und somit würde die Forschungspraxis (und auch die Prüfung der Erforderlichkeit) dem gesetzlichen Imperativ zur möglichen Leidminimierung als notwendige Bedingung für die ethische Vertretbarkeit eines Tierversuchs nur unzureichend gerecht.¹⁰ Um Forschung, Lehre und die praktische Implementierung der 3R zu fördern, wurden daher in den letzten Jahrzehnten europaweit in verschiedenen Ländern sogenannte 3R-Zentren aufgebaut (z. B. NC3Rs in England, 3RCC in der Schweiz etc.).

Des Weiteren ist es wichtig anzumerken, dass es sich bei der Prüfung auf Erforderlichkeit im Genehmigungsverfahren ausschließlich um eine Prüfung der *wissenschaftlichen* Notwendigkeit handelt, welche von einer *ethischen* Notwendigkeit abzugrenzen ist. Nur weil sich ein Versuch und ein damit verbundenes Ziel ausschließlich durch die Verwendung von Tie-

9 Laut der Veterinärmedizinerin Petra Seebeck von der Universität Zürich können gängige Operationspraktiken noch stark verbessert werden (Petra Seebeck [Interview with] 2020).

10 Hier muss natürlich angemerkt werden, dass die Prüfung der Erforderlichkeit als *ex ante* Prüfung nur prüfen kann, wie ein Versuch und dessen Durchführung geplant ist, aber keine Kontrolle hat, wie der Versuch schlussendlich im Labor durchgeführt wird.

ren erreichen lässt und deshalb der Tierversuch wissenschaftlich notwendig ist, ergibt sich daraus nicht zwangsläufig, dass dieser Versuch auch ethisch geboten wäre. Dementsprechend ist die „ethische Reichweite“ des 3R-Prinzips (und seiner innewohnenden *animal welfare*-Position) sehr begrenzt in dem Sinne, dass tierliches Leid nur insoweit berücksichtigt wird, als es die wissenschaftlichen Interessen nicht beeinträchtigt. Laut Richtlinie ist diese Prüfung der wissenschaftlichen Erforderlichkeit denn auch nur eine notwendige, aber keine hinreichende Bedingung für die ethische Vertretbarkeit eines Tierversuchs. Diese hängt laut Gesetz zum einen auch von der Eignung und zum anderen vom Verhältnis zwischen erwartetem Erkenntnisgewinn/Nutzen und den Belastungen der Tiere ab. Diese finale Evaluierung der „ethischen Vertretbarkeit“ eines Tierversuchs ist gesetzlich durch die Prüfung der Verhältnismäßigkeit geregelt, welche im nächsten Abschnitt behandelt wird.

2.3 Prüfung der Verhältnismäßigkeit

Damit ein Tierversuch verhältnismäßig (und somit nach Gesetz „ethisch vertretbar“) ist, verlangt Artikel 38/d der Richtlinie

„eine Schaden-Nutzen-Analyse des Projektes, in deren Rahmen bewertet wird, ob die Schäden für die Tiere in Form von Leiden, Schmerzen und Ängsten unter Berücksichtigung ethischer Erwägungen durch das erwartete Ergebnis gerechtfertigt sind und letztlich Menschen, Tieren oder der Umwelt zugutekommen können“.

Diese Prüfung entspricht also einer Güterabwägung zwischen den tierlichen Interessen der Leid- und Schmerzvermeidung und dem menschlichen Interesse nach Erkenntnisgewinn. Die Bedeutung einer solchen, gesetzlich geforderten Güterabwägung ist signifikant, zeigt sie doch eine Abkehr von einem Forschungsparadigma an, laut welchem unerlässliche und zweckgeeignete Versuche und Versuchsdesigns nicht alleine aufgrund ihrer wissenschaftlichen Notwendigkeit ethisch vertretbar sind (Klaus Peter Rippe 2009). Das widerspiegelt die Möglichkeit, dass ein Versuch, obwohl nach wissenschaftlichen Kriterien exzellent und notwendig, aufgrund hoher Belastungen für die Tiere unverhältnismäßig und unzulässig sein kann. Das Prinzip der Güterabwägung steht somit in starkem Kontrast zum 3R-Prinzip, welchem eine klare Vorrangregelung der menschlichen Interessen gegenüber tierlichen Interessen innewohnt. Demnach sind laut 3R-Prinzip Tierversuche ethisch gerechtfertigt, wenn sie wissenschaftlich erforderlich sind und die Belastungen der Tiere minimal gehalten wer-

den. Die Güterabwägung bedeutet also eine Abkehr von einer klaren Vorrangregelung hin zu einer rational begründeten Vorzugswahl „jener Handlung, welche die bestmögliche Differenz zwischen Vor- und Nachteilen, d. h. den größten Nettonutzen, erbringt“ (Klaus Peter Rippe 2009, 4). Eine solche Güterabwägung entspringt der utilitaristischen Moralphilosophie, nach der die Güte einer Handlung davon abhängt, inwieweit sie den Nutzen (d. h. das Gute, Wohlergehen, Freude) oder das größtmögliche Glück für die größtmögliche Zahl fördert. Das utilitaristische Paradigma versucht, das Ausmaß von Nutzen (oder die Zufriedenstellung von Interessen) zu maximieren und das Ausmaß an Schmerz (oder der Frustration von Interessen) zu minimieren (Kerstin Weich 2018). Für den Tierversuch bedeutet dies nun, dass die Frustration der tierlichen Interessen nur dann gerechtfertigt ist, wenn damit andere, überwiegende Interessen gefördert werden. Dies verlangt logischerweise eine Klassifizierung von Belastungen und Nutzen sowie deren Abwägung. Im Folgenden sollen zwei Probleme skizziert werden: zum einen die Frage, was prospektiv als Nutzen zu werten und zu gewichten ist; zum anderen das Problem der begrenzten Reichweite „ethischer Erwägungen“ innerhalb der Schaden-Nutzen-Analyse (als gesetzlich vorgeschriebener Prüfschritt).

Schwierigkeiten bei der Bestimmung von Nutzen

Die Güterabwägung in Form einer Schaden-Nutzen-Analyse (SNA) liegt in der Verantwortung der genehmigenden Behörden, die in ihrer Aufgabe oft durch Ethikkommissionen unterstützt werden. Die Zusammensetzung dieser Kommissionen kann sehr heterogen sein und beinhaltet normalerweise Experten aus den Gebieten der Forschung, Veterinärmedizin, Statistik, Rechtswissenschaften und Ethik sowie oft auch Tierschutzvertreter und „Laien“.

Die Richtlinie schweigt sich darüber aus, was Nutzen genau ist und wie dieser gewogen werden soll. Deshalb sind die Bestimmung und die Klassifizierung von Nutzen keineswegs trivial und werfen auf ethischer wie auf praktischer Ebene verschiedene Fragen auf. Ist z. B. die Entwicklung eines neuen Impfstoffs höher zu gewichten als Grundlagenwissen zu biologischen Prozessen? Gibt es eine Hierarchie zwischen verschiedenen Krankheiten? Wenn es eine Hierarchie zwischen verschiedenen Arten von Nutzen gibt, wie wird diese begründet? Oder sind „alle“ Arten von Nutzen als gleichwertig zu gewichten?

Als Hilfestellung für die genehmigenden Behörden wurde eine Expertengruppe von der Europäischen Kommission eingerichtet, um sich dieser

Fragen anzunehmen. In ihrem Bericht empfiehlt die Expertengruppe, generell folgende Fragen zu prüfen (European Commission/Expert Working Group 2013, 21): Was wird der erwartete Nutzen des Projekts sein? Wer wird vom Projekt profitieren? Wie wird man profitieren können? Wann (wo möglich) wird der Nutzen realisiert? Laut Expertengruppe soll zuerst der kurzfristig zu erwartende Nutzen beurteilt werden. Auf die Grundlagenforschung bezogen meint der Bericht damit, dass der Erkenntnisgewinn und dessen Wichtigkeit beurteilt werden sollen. Laut Bericht sind Projekte, die einen kurzfristigen sozialen Nutzen versprechen, wie z. B. die Entwicklung eines neuen Impfstoffs, leichter zu beurteilen als Grundlagenforschung. Es soll aber auch in der Grundlagenforschung, wenn möglich, der Erkenntnisgewinn als „primäres Ergebnis“ an ein weiterreichendes Ziel gekoppelt werden (z. B. Langzeit-Nutzen und translationales Potential für die Klinik), auch wenn dieses weiter in der Zukunft liegt und schwieriger abzuschätzen ist. Die Expertengruppe wünscht sich diesbezüglich Angaben zum Ausmaß „der Verbesserung“ (z. B. für Mensch, Tier und Umwelt; Anzahl, Qualität etc.), welche der erwartete Nutzen liefert. Ein höherer „Nutzenwert“ wird diesbezüglich je nach Ausmaß der Verbesserung für die menschliche Gesundheit (und je nach Anzahl der betroffenen Menschen) zugesprochen. Zentrales Augenmerk liegt hier auch auf einer Abschätzung der Wahrscheinlichkeit, mit welcher der erwartete Nutzen tatsächlich eintreffen wird.

Der Bericht spricht also dem Erkenntnisgewinn zwar erhebliche Signifikanz zu und erwähnt auch explizit, dass es nicht möglich ist, verschiedene Arten von Nutzen in eine klare Hierarchie einzuteilen (European Commission/Expert Working Group 2013, 22). Gleichzeitig wird aber auch ganz klar eine Hierarchie etabliert, in welcher sozialer Nutzen (z. B. Entwicklung neuer Medikamente oder Technologien, die der Gesundheit des Menschen förderlich sind) als „gewichtiger“ beurteilt wird als Erkenntnisgewinn.

Diese Priorisierung von sozialem Nutzen und dessen Quantifizierung (z. B. Ausmaß der Verbesserung und Anzahl der Begünstigten) gegenüber Erkenntnisgewinn in einer prospektiven Prüfung wurde von Forschungsseite aus praktischen Gründen stark kritisiert. Es wurde die kurze (zeitliche wie räumliche) Kausalkette zwischen einem Tierversuchsprojekt und der Entwicklung und Fertigstellung von neuen Therapien oder Technologien, auf welcher eine *ex ante* Beurteilung von praktischem Nutzen zwangsläufig beruht, verworfen. Es seien nicht einzelne Projekte, die innert kürzester und absehbarer Zeit praktischen Nutzen abwerfen würden, sondern eine Vielzahl (oft hunderte) von Projekten über einen großen Zeithorizont (oft mehrere Jahrzehnte) hinweg, die zusammen (ähnlich dem Bild eines Man-

dalas) zur Entwicklung und Fertigstellung sozialen Nutzens führe. Dementsprechend beruhe die Vorstellung, auf der Projektebene Anträge prospektiv auf ihr Potential und die Wahrscheinlichkeit, sozialen Nutzen zu generieren, zu beurteilen, auf einer praktischen Unmöglichkeit und sei dementsprechend eine fehlgeleitete Idee.¹¹ Vielmehr solle auf der Stufe der prospektiven Prüfung das Augenmerk auf dem erwarteten Erkenntnisgewinn liegen als zentraler Prüfstein für die Genehmigung. Der soziale Nutzen solle und könne erst retrospektiv beurteilt werden. Dementsprechend müsse sich aus praktischen Gründen in einer *ex ante* Beurteilung die Nutzenbeurteilung innerhalb der Prüfung der Verhältnismäßigkeit auf eine Abwägung des erwarteten Erkenntnisgewinns (mit ungewissem sozialem Nutzen) beschränken. Eine Hierarchisierung verschiedener Arten von Nutzen scheint nicht nur aus praktischen, sondern auch aus juristischen Gründen nicht angezeigt. Das Obergerverwaltungsgericht Bremen spricht in einem Gerichtsurteil der genehmigenden Behörde die Autorität ab, eine Hierarchie zwischen den verschiedenen gesetzlich vorgesehenen Zwecken (definiert in Artikel 5 der Richtlinie¹²) und deren Nutzen festzulegen (Oberverwaltungsgericht Bremen 2012).¹³

Grenzen „ethischer Erwägungen“ innerhalb der Prüfung auf Verhältnismäßigkeit

Wie erwähnt folgt laut Richtlinie aus der Zweckeignung und wissenschaftlichen Notwendigkeit eines Tierversuchs nicht zwingend, dass dieser auch ethisch gerechtfertigt ist. Diesbezüglich sieht Artikel 38/d der Richtlinie eine finale Güterabwägung vor, welche die Belastungen für das Tier gegen den erwarteten Nutzen „unter Berücksichtigung ethischer Erwägungen“ abwägen soll. Verlangt wird nämlich

-
- 11 Herwig Grimm et al. (2017) veranschaulichen dieses Problem anhand logischer, methodischer und praktischer Hindernisse der prospektiven (*ex ante*) Nutzenbeurteilung.
 - 12 Grundlagenforschung, translationale oder angewandte Forschung, regulatorische Versuche, Schutz der Umwelt im Interesse der Gesundheit und Wohlergehen von Mensch und Tier, Forschung zur Erhaltung der Arten, Ausbildung und forensische Untersuchungen.
 - 13 Die deutsche Gesetzgebung scheint in diesem Aspekt der österreichischen und schweizerischen Gesetzgebung sehr ähnlich, weshalb es in den Augen des Autors (der aber kein Jurist ist) plausibel erscheint, dass dieses Gerichtsurteil auch für die Schweiz und für Österreich von Belang ist.

„eine Schaden-Nutzen-Analyse des Projektes, in deren Rahmen bewertet wird, ob die Schäden für die Tiere in Form von Leiden, Schmerzen und Ängsten unter Berücksichtigung ethischer Erwägungen durch das erwartete Ergebnis gerechtfertigt sind und letztlich Menschen, Tieren oder der Umwelt zugutekommen können“.

Weil sich die Richtlinie darüber ausschweigt, was unter „Berücksichtigung ethischer Erwägungen“ genau zu verstehen ist, ist es unklar, was in die Prüfung der Verhältnismäßigkeit einfließen soll. Wie oben beschrieben, kann aus juristischen und praktischen Gründen damit nicht die Wertung/Hierarchisierung verschiedener Arten von Nutzen gemeint sein. Laut Herwig Grimm (2015) und Kerstin Weich (2018) ist die „Berücksichtigung ethischer Erwägungen“ auch nicht mit dem Einbezug öffentlicher Bedenken oder mit moralphilosophischer Kritik des Gesetzes gleichzusetzen. Die Forderung, „ethische Erwägungen“ in die Güterabwägung einzubeziehen, nehme eine Zwitterstellung zwischen den Sphären von Ethik und Recht ein (Kerstin Weich 2018) und bedeute einen Prüfschritt „am Schnittpunkt zwischen ethischer Verantwortung und rechtlicher Verpflichtung“ (Regina Binder/Herwig Grimm 2013, 56). Das Legalitätsprinzip fordere, dass in einem Verfassungsstaat Behörden nur aufgrund geltenden Rechts Entscheidungen treffen dürfen. Da es sich bei der Genehmigung von Tierversuchen um einen gesetzlich festgelegten Auftrag handelt, könne Ethik hier nur innerhalb der Grenzen des geltenden Gesetzes wirksam sein. Dementsprechend seien auch die Behördenvertreter als ausführende Instanz an das geltende Gesetz gebunden. Deshalb gehe es also nicht darum, z. B. die Idee einer Güterabwägung in Frage zu stellen, die Vor- und Nachteile von verschiedenen Moraltheorien zu diskutieren oder auf die aktuelle öffentliche Meinung Rücksicht zu nehmen. Der Ethik komme in diesem Kontext primär die Rolle zu, die normativen Prinzipien des Gesetzes explizit zu machen, d. h. die normative Infrastruktur der Richtlinie herauszuarbeiten, damit Behördenvertreter wissen, was sie in ihre Beurteilung überhaupt einfließen lassen dürfen (und müssen). In nuce basiert die Richtlinie primär auf einem pathozentrischen Utilitarismus. Pathozentrisch, weil das Hauptaugenmerk auf der Vermeidung von erlebnisrelevanten, negativen, subjektiven Empfindungen von empfindungsfähigen Tieren – wie z. B. Schmerzen, Leiden, Schäden, Angst – liegt. Utilitaristisch, weil die Abwägung verschiedener rechtlich geschützter Interessen angezeigt ist.¹⁴ Weil eben schon im Voraus festgelegt ist, dass die Beurteilung in diesem normativen

14 Zudem basiert die Richtlinie auf einer anthropozentrischen Position. Dies ist zwar nirgends in der Richtlinie explizit erwähnt, wird aber evident durch die Tat-

Rahmen zu erfolgen hat, ist offensichtlich, warum mit „ethischen Erwägungen“ kein Abwägen der Vor- und Nachteile verschiedener Moraltheorien gemeint sein kann. Als Teil eines gesetzlich vorgeschriebenen Verfahrens kann in diesem Zusammenhang unter „Ethik“ oder „ethischen Erwägungen“ nur eine ganz spezifische und enggeführte Idee von „Ethik“ verstanden werden, welche eben das Abwägen innerhalb des vorgegebenen normativen Rahmens bedeutet. Mit Bezug auf die beschriebene Problematik der Nutzenbestimmung bedeutet das, dass sich „ethische Erwägungen“ im Behördenverfahren darauf beschränken, innerhalb eines pathozentrischen und utilitaristischen Rahmens den erwarteten Erkenntnisgewinn (unabhängig des potentiellen sozialen Nutzens) und die Belastungen am Tier auf ihre Verhältnismäßigkeit zu prüfen, d. h. abzuwägen, wie relevant ein Forschungsvorhaben für ein Forschungsfeld ist und inwieweit die Belastungen durch den Erkenntnisgewinn gerechtfertigt sind. Als solches kann (soll und darf) das Behördenverfahren mit seinen gesetzlichen Beschränkungen eine ethische Reflexion im eigentlichen Sinne nicht ersetzen (und soll auch nicht als solche missverstanden werden). Vielmehr scheint die Frage angebracht, inwieweit man in einem solchen Kontext überhaupt von „Ethik“ reden kann oder soll. Durch die begrenzte Reichweite der behördlichen Prüfung macht diese die Notwendigkeit einer ethischen Reflexion sogar evident (Herwig Grimm 2015).

Eine kritische Auseinandersetzung mit Tierversuchen und den (keinesfalls selbst-evidenten) normativen Annahmen, auf welchen die EU-Richtlinie 2010/63/EU beruht, kann aber nur, losgelöst von Gesetz und Genehmigungsverfahren, im akademischen sowie im gesellschaftlichen Diskurs stattfinden. Hier kann (und muss) Ethik nun die Frage stellen, ob Tierversuche sich durch die utilitaristische Position der Richtlinie tatsächlich rechtfertigen lassen und „ob eine Gesellschaft wünschenswert ist, in der das (moralisch) Gute mit dem Nützlichen“ assoziiert wird (Kerstin Weich 2018, 287). Hier gilt es zu hinterfragen, was die Defizite und Grenzen der

sache, dass Versuche an Tieren erlaubt, Versuche an Menschen aber verboten sind und somit menschliche Interessen mehr wiegen als die von Tieren. Auch enthält die Richtlinie eine (deontologische) Schmerzobergrenze, welche unabhängig vom Nutzen nicht überschritten werden darf. Gemessen an den in Anhang VIII aufgeführten Beispielen für Eingriffe, die als „schwere Belastung“ eingestuft werden (aber noch unter der Schmerzobergrenze sind), kann angenommen werden, dass die große Mehrheit an zu beurteilenden Versuchen diese (deontologische) Belastungsobergrenze nicht erreicht und diese daher nur in einer Minderheit der Fälle in die Entscheidung mit einbezogen wird.

normativen Konzepte des Gesetzes sind: Von welchen Annahmen geht es aus, was sind die Auswirkungen und welche Alternativen gibt es?

3. Zusammenfassung

Ziel dieses Aufsatzes war es, die Prüfung auf wissenschaftliche Erforderlichkeit, Eignung und Verhältnismäßigkeit als drei der Eckpfeiler des Genehmigungsverfahrens zu beschreiben, welche laut Gesetz erfüllt sein müssen, damit ein Tierversuch ethisch vertretbar ist. Diesen Konzepten liegt die Annahme zugrunde, dass nur Forschung ethisch vertretbar ist, welche valide und belastbare Resultate liefert, Tiere nur dann verwendet werden dürfen, wenn es wissenschaftlich notwendig ist und wenn die Belastungen für das Tier auf ein Minimum reduziert werden. Zudem muss der erwartete Nutzen unter „Berücksichtigung ethischer Erwägungen“ die Belastungen aufwiegen/rechtfertigen. Wie dieser Aufsatz versucht hat zu zeigen, gibt es für alle drei Prüfschritte Probleme auf praktischer Ebene und die „ethische Reichweite“ jedes Prüfschritts ist sehr begrenzt. Die mangelnde Qualität der Forschung (und die ungenügende Effektivität der Prüfung auf Eignung) ist kritisiert worden und bedarf der Verbesserung, um ihrem Anspruch gerecht zu werden. Obwohl die 3R weithin als „ethisches Leitprinzip“ in der Forschung anerkannt sind, wurde skizziert, dass die Prüfung der Erforderlichkeit ausschließlich eine Prüfung auf wissenschaftliche Notwendigkeit darstellt (von welcher sich noch keine ethische Rechtfertigung ableiten lässt, einen Tierversuch durchzuführen). Zudem lassen die erwähnten Beispiele vermuten, dass die vorhandenen Möglichkeiten für *Replacement*, *Refinement* und *Reduction* auf praktischer Ebene nicht immer ausgeschöpft werden. Die Prüfung der Verhältnismäßigkeit mittels Schaden-Nutzen-Analyse ist das Instrument der Wahl der Richtlinie zur finalen Entscheidung über die „ethische Vertretbarkeit“ von Tierversuchen. Hier reduzieren praktische und juristische Gründe die Prüfung auf eine Beurteilung der Verhältnismäßigkeit von erwartetem Erkenntnisgewinn und Belastungen am Tier innerhalb eines vorgegebenen pathozentrischen und utilitaristischen rechtlichen Rahmens. Diesbezüglich kann die gesetzliche Forderung (und ihr Anspruch) der „Berücksichtigung ethischer Erwägungen“ eine kritische Reflexion der Praxis des Tierversuchs nicht ersetzen und darf nicht als solche missverstanden werden (zumindest nicht mit einer Reflexion, die über das bestehende Recht hinausgeht). Eine kritische Auseinandersetzung mit Tierversuchen und den (keinesfalls selbst-evidenten) normativen Annahmen, auf welchen die EU-Richtlinie 2010/63/EU

beruht, kann nur losgelöst von Gesetz und Genehmigungsverfahren im akademischen sowie im gesellschaftlichen Diskurs stattfinden.¹⁵

Literatur

- Baker, David/Lidster, Katie/Sottomayor, Ana/Amor, Sandra 2014, Two Years Later: Journals Are Not Yet Enforcing the ARRIVE Guidelines on Reporting Standards for Pre-Clinical Animal Studies, in: PLoS Biology 12 (1): e1001756; DOI: org/10.1371/journal.pbio.1001756.
- Binder, Regina/Grimm, Herwig 2013, Die Schaden-Nutzen-Analyse, in: Binder, Regina/Alzmann, Norbert/Grimm, Herwig (Hg.) 2013, Wissenschaftliche Verantwortung im Tierversuch. Ein Handbuch für die Praxis, Baden-Baden, 55-67.
- Button, Katherine S./Ioannidis, John P. A./Mokrysz, Claire/Nosek, Brian A./Flint, Jonathan et al. 2013, Power Failure: Why Small Sample Size Undermines the Reliability of Neuroscience, in: Nature Reviews Neuroscience 14 (5), 365-376; DOI: 10.1038/nrn3475.
- Europäisches Parlament/Rat der Europäischen Union 2010, Richtlinie 2010/63/EU zum Schutz der für wissenschaftliche Zwecke verwendeten Tiere (22.09.2010), in: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2010:276:0033:0079:de:PDF> (abgerufen am 10.06.2020).
- European Commission/Expert Working Group 2013, National Competent Authorities for the implementation of Directive 2010/63/EU on the protection of animals used for scientific purposes. Working document on Project Evaluation and Retrospective Assessment Brussels (18.-19.09.2013), in: https://ec.europa.eu/environment/chemicals/lab_animals/pdf/Endorsed_PE-RA.pdf (abgerufen am 10.06.2020).
- Francione, Gary L. 1996, Animal Rights and Animal Welfare, in: Rutgers Law Review 48, 397-469.
- Grimm, Herwig 2015, Ethics within Legal Limits: Harm-Benefit Analysis According to the Directive 2010/63/EU, in: Dumitras, Diana Elena/Jitea, Ionel Mugurel/Aerts, Stef (Hg.) 2015, Know your food. Food ethics and innovation, Wageningen, 42-47; DOI: org/10.3920/978-90-8686-813-1_5.
- Grimm, Herwig/Eggel, Matthias/Deplazes-Zemp, Anna/Biller-Adorno, Nikola 2017, The Road to Hell Is Paved with Good Intentions: Why Harm-Benefit Analysis and Its Emphasis on Practical Benefit Jeopardizes the Credibility of Research, in: Animals (Basel) 7, art. 70; DOI: 10.3390/ani7090070.

15 Dank gilt Samuel Camenzind und Paulin Jirkof für die kritische Durchsicht des Aufsatzes und für hilfreiche Kommentare sowie Hille Haker und Martin M. Lintner für konstruktive Rückmeldungen. Weiterer Dank gilt Dagmar Chastonay für das Korrekturlesen des Manuskripts.

- Hair, Kaitlyn/Macleod, Malcom R./Sena, Emily S./IICARus Collaboration 2019, A randomised controlled trial of an Intervention to Improve Compliance with the ARRIVE guidelines (IICARus), in: Research Integrity and Peer Review 4, art. 12; DOI: 10.1186/s41073-019-0069-3.
- Henderson, Lindsay J./Smulders, Tom V./Roughan, Johnny V. 2020, Identifying obstacles preventing the uptake of tunnel handling methods for laboratory mice: An international thematic survey, in: PLoS ONE 15 (4): e0231454; DOI: org/10.1371/journal.pone.0231454.
- Herrmann, Kathrin/Flecknell, Paul 2018 a, Severity Classification of Surgical Procedures and Application of Health Monitoring Strategies in Animal Research Proposals: A Retrospective Review, in: Alternatives to Laboratory Animals 46 (5), 273-289.
- Herrmann, Kathrin/Flecknell, Paul 2018 b, The Application of Humane Endpoints and Humane Killing Methods in Animal Research Proposals: A Retrospective Review, in: Alternatives to Laboratory Animals 46 (6), 317-333; DOI: org/10.1177/026119291804600606.
- Karp, Natasha A./Speak, Anneliese O./White, Jacqueline K./Adams, David J./Hrabé de Angelis, Martin 2014, Impact of Temporal Variation on Design and Analysis of Mouse Knockout Phenotyping Studies, in: PLoS ONE 9 (10): e11239; DOI: doi.org/10.1371/journal.pone.0111239.
- Kilkenny, Carol/Browne, William J./Cuthill, Innes C./Emerson, Michael/Altman, Douglas G. 2010, Improving Bioscience Research Reporting: The ARRIVE Guidelines for Reporting Animal Research, in: PLoS Biology 8 (6): e1000412; 10.1371/journal.pbio.1000412.
- Knight, Andrew 2008, 127 Million Non-human Vertebrates Used Worldwide for Scientific Purposes in 2005, in: Alternatives to Laboratory Animals 36 (5), 494-496.
- Kola, Ismail/Landis, John 2004, Can the pharmaceutical industry reduce attrition rates?, in: Nature Reviews Drug Discovery 3, 711-716; doi: 10.1038/nrd1470.
- Mazhary, Hibba/Hawkins, Penny 2019, Applying the 3Rs: A Case Study on Evidence and Perceptions Relating to Rat Cage Height in the UK, in: Animals (Basel) 9 (12), art. 1104; DOI: 10.3390/ani9121104.
- O'Collins, Victoria E./Macleod, Malcom R./Donnan, Geoffrey A./Horky, Laura L./van der Worp, H. Bart et al. 2006, 1,026 Experimental Treatments in Acute Stroke, in: Annals of Neurology 59 (3), 467-477; DOI: 10.1002/ana.20741.
- Oberverwaltungsgericht Bremen 2012, Entscheidung „Tierversuche an der Universität Bremen“ (11.12.2012), in: <http://www.oberverwaltungsgericht.bremen.de/sixcms/detail.php?gclid=bremer72.c.11099.de&asl=bremer72.c.11265> (abgerufen am 10.06.2020).
- Paylor, Richard 2009, Questioning standardization in science, in: Nature Methods 6, 253-254; DOI:10.1038/nmeth0409-253.
- Reichlin, Thomas S./Vogt, Lucile/Würbel, Hanno 2016, The Researchers' View of Scientific Rigor – Survey on the Conduct and Reporting of *In Vivo* Research, in: PLoS ONE 11 (12): e0165999; DOI: org/10.1371/journal.pone.0165999.

- Richter, S. Helene/Garner, Joseph P./Würbel, Hanno 2009, Environmental Standardization: Cure or Cause of Poor Reproducibility in Animal Experiments?, in: *Nature Methods* 6, 257-261; DOI: 10.1038/nmeth.1312.
- Richter, S. Helene/Garner, Joseph P./Auer, Corinna/Kunert, Joachim/Würbel, Hanno 2010, Systematic Variation Improves Reproducibility of Animal Experiments, in: *Nature Methods* 7, 167-168; DOI: 10.1038/nmeth0310-167.
- Richter, S. Helene/Garner, Joseph P./Zipser, Benjamin/Lewejohann, Lars/Sachser, Norbert et al. 2011, Effect of Population Heterogenization on the Reproducibility of Mouse Behavior: A Multi-Laboratory Study, in: *PLoS ONE* 6 (1): e16461; DOI: org/10.1371/journal.pone.0016461.
- Rippe, Klaus Peter 2009, Güterabwägungen im Tierversuchsbereich. Anmerkungen zu einem ethischen Paradigmenwechsel, in: *ALTEXethik* 26 (1), 3-10.
- Russell, William/Burch, Rex 1959, *The Principles of Humane Experimental Technique*, London.
- Seebeck, Petra (Interview with) 2020, Rodents have a right for best surgical practice, in: *Zurich Center for Integrative Human Physiology News*, January, 1-2, in: https://www.zihp.uzh.ch/static/cms/newsletter_cms/ZIHPNews_Jan20_web.pdf (abgerufen am 10.06.2020).
- Sena, Emily S./van der Worp, H. Bart/Bath, Philip M. W./Howells, David W./Macleod, Malcolm R. 2010, Publication Bias in Reports of Animal Stroke Studies Leads to Major Overstatement of Efficacy, in: *PLoS Biology* 8 (3): e1000344; DOI: org/10.1371/journal.pbio.1000344.
- Smith, Adrian/Clutton, Eddie R./Lilley, Elliot/Hansen, Kristine E. Aa./Brattelid, Trond 2018, PREPARE: guidelines for planning animal research and testing, in: *Laboratory Animals* 52 (2), 135-141: e002367721772482; DOI: 10.1177/0023677217724823.
- Strech, Daniel/Dirnagl, Ulrich 2019, 3Rs missing: animal research without scientific value is unethical, in: *BMJ Open Science* 3: e000035; DOI: 10.1136/bmjos-2018-000048.
- Swiss Academies of Arts and Sciences 2017, Weighing of interests for proposed animal experiments. Guidance for applicants = Swiss Academies Communications 12 (3).
- Vogt, Lucile/Reichlin, Thomas S./Nathues, Christina/Würbel, Hanno 2016, Authorization of Animal Experiments Is Based on Confidence Rather than Evidence of Scientific Rigor, in: *PloS Biology* 14 (12): e2000598; DOI: 10.1371/journal.pbio.2000598.
- Weich, Kerstin 2018, Privater oder öffentlicher Vernunftgebrauch? Zur Prüfung auf ethische Vertretbarkeit von Tierversuchsvorhaben, in: *Berliner und Münchener Tierärztliche Wochenschrift* 131 (7/8), 284-291; DOI: 10.2376/0005-9366-17099.
- Wodarski, Rachel/Delaney, Ada/Ultenius, Camilla/Morland, Rosie/Andrews, Nick et al. 2016, Cross-centre Replication of Suppressed Burrowing Behaviour as an Ethologically Relevant Pain Outcome Measure in the Rat: A Prospective Multi-centre Study, in: *Pain* 157 (10), 2350-2365; DOI: 10.1097/j.pain.000000000000057.

- van der Worp, H. Bart/Howells, David W./Sena, Emily S./Porritt, Michelle J./Rewell, Sarah et al. 2010, Can Animal Models of Disease Reliably Inform Human Studies?, in: PLoS Medicine 7 (3): e1000245; DOI: 10.1371/journal.pmed.1000245.
- Würbel, Hanno 2000, Behaviour and the standardization fallacy. Nature Genetics 26 (3), 263; DOI: org/10.1038/81541.
- Würbel, Hanno 2001, Ideal Homes? Housing Effects on Rodent Brain and Behaviour, in: Trends in Neurosciences 24 (4), 207-211; DOI: 10.1016/s0166-2236(00)01718-5.
- Würbel, Hanno 2017, More Than 3Rs: The Importance of Scientific Validity for Harm-Benefit Analysis of Animal Research, in: LabAnimal 46 (4), 164-166; DOI: 10.1038/labani.1220.