

Tierversuche in der Forschung: Das 3R-Prinzip und die Aussage- kraft wissenschaftlicher Forschung

**Animal Experimentation in Research:
The 3Rs Principle and the Validity of
Scientific Research**



**Handreichung der Ständigen Senatskommission
für tierexperimentelle Forschung der DFG zur
Planung und Beschreibung tierexperimenteller
Forschungsprojekte**

**Guidelines of the Permanent Senate Commission on Animal
Protection and Experimentation of the DFG for the Design
and Description of Animal Experimental Research Projects**

Stand: Juli 2019

Ansprechpartnerin:

Dr. Sonja Ihle

Gruppe Lebenswissenschaften 1

Tel.: 0228/885-2362

E-Mail: sonja.ihle@dfg.de

Inhaltsverzeichnis

1	Zusammenfassung	3
2	Zusammenhang zwischen Tierschutz in der Forschung und wissenschaftlicher Qualität	5
3	Mehr als 3R – 3R-Prinzip und wissenschaftliche Aussagekraft	7
3.1	Legitimation von tierexperimenteller Forschung durch Güterabwägung	7
3.2	Das 3R-Prinzip als normatives Kriterium für den Tierschutz in der Forschung.....	8
3.3	Das 3R-Prinzip im Kontext von Validität und Replizierbarkeit	8
3.4	Aspekte der wissenschaftlichen Aussagekraft von Tierversuchen	9
4	Spannungsfelder zwischen der wissenschaftlichen Validität und den 3R	12
4.1	Wahl der Tierart	12
4.2	Festlegung der Tierzahl	14
4.3	Standardisierung	15
4.4	Pilot- und Explorationsversuche	16
4.5	Wiederholungsversuche	17
5	Anhang	19
6	Mitglieder der Ständigen Senatskommission für tierexperimentelle Forschung der Deutschen Forschungsgemeinschaft	20

1 Zusammenfassung

Das im Grundgesetz verankerte Recht auf Freiheit der Forschung und das im Grundgesetz benannte Staatsziel Tierschutz sind zwei bedeutende Güter, die für die Belange der tierexperimentellen Forschung eine hohe Relevanz haben.

Die Verwendung von Tieren zu wissenschaftlichen Zwecken ist nur dann ethisch zu rechtfertigen, wenn ein Tierversuch „unerlässlich“ ist und der zu erwartende wissenschaftliche Erkenntnisgewinn die Belastung der Tiere im Versuch rechtfertigt.

Wesentlicher ethischer Leitgedanke in der tierexperimentellen Forschung ist das 3R-Prinzip (Replace, Reduce, Refine)¹: Tierversuche dürfen nur dann durchgeführt werden, wenn zur Aufklärung der wissenschaftlichen Fragestellung keine anderen geeigneten Methoden zur Verfügung stehen und die Anzahl und Belastung der eingesetzten Tiere auf das unerlässliche Maß beschränkt werden. Die konsequente Umsetzung des 3R-Prinzips ist im ureigenen Interesse der Wissenschaft, da eine Beeinträchtigung des Wohlergehens der Versuchstiere auch die Aussagekraft der Forschungsergebnisse beeinflussen kann. Der Anspruch, ein Höchstmaß an wissenschaftlicher Validität und Replizierbarkeit der Ergebnisse zu erzielen, bei gleichzeitiger Berücksichtigung von tierschutzrelevanten Maßnahmen, muss die Grundlage jeder tierexperimentellen Versuchsplanung sein.

In der Praxis kann es zwischen den Maßnahmen zur Umsetzung des Tierschutzes in der Forschung und den Anforderungen an die wissenschaftliche Validität zu Spannungen kommen. Aus der Wechselwirkung der beiden Aspekte ergibt sich die Notwendigkeit, Maßnahmen zur Umsetzung des 3R-Prinzips nicht separat zu betrachten, sondern als integralen Bestandteil wissenschaftlicher Projektplanung zu verankern und im Rahmen der Beschreibung von Forschungsprojekten zu berücksichtigen.

1 Russell, W. M. S., Burch, R. L. (1959): The principles of humane experimental technique. London: Methuen.

Die Ständige Senatskommission für tierexperimentelle Forschung der DFG möchte mit dieser Handreichung einen Beitrag zur Debatte über Qualität in der biomedizinischen Forschung liefern und die spezifischen Anforderungen an die Durchführung von Tierversuchen herausarbeiten. Die Handreichung bietet darüber hinaus Hilfestellung zur validen Planung und Beschreibung von tierexperimentellen Forschungsprojekten.

2 Zusammenhang zwischen Tierschutz in der Forschung und wissenschaftlicher Qualität

Wissenschaft basiert auf dem Anspruch, ein Höchstmaß an Qualität und Aussagekraft der Forschungsergebnisse anzustreben. Im Rahmen der Diskussionen über eine „Replikationskrise“ in der Wissenschaft^{2,3} hat die DFG in ihren Stellungnahmen zur Replizierbarkeit von Forschungsergebnissen die Sicherung der Qualität von Forschung als grundlegenden Anspruch an die Wissenschaft hervorgehoben und stößt einen Diskurs zur fachspezifischen Reflexion über Qualität und Replizierbarkeit sowie verschiedene Aspekte der Qualitätssicherung an.^{4,5}

In der tierexperimentellen Forschung besteht die Notwendigkeit, die wissenschaftliche Aussagekraft von Ergebnissen und zugleich einen adäquaten und verantwortlichen Umgang mit Versuchstieren zu gewährleisten. Hohe Tierschutzstandards stellen dabei eine Voraussetzung für qualitativ hochwertige Forschung dar, und umgekehrt sind Tierversuche nur zu rechtfertigen, wenn die Qualität der Forschung und damit die Aussagekraft von Forschungsergebnissen sichergestellt sind.

Aufgrund der grundsätzlichen Schutzwürdigkeit von Tieren tragen Forschende eine besondere moralische Verantwortung bei der Verwendung von Tieren. Eine kriteriale Orientierung für Maßnahmen zum Tierschutz beschreibt das 3R-Prinzip, nämlich die Verfeinerung (refinement), Verringerung (reduction)

-
- 2 Ioannidis, J. (2005): Why most published research findings are false. *PLoS Medicine*, 2(8): 696–701.
 - 3 McLeod, M. R. et al. (2014): Increasing value, reducing waste. *The Lancet*, 383(9912): 101–104.
 - 4 Deutsche Forschungsgemeinschaft (2017): Replizierbarkeit von Forschungsergebnissen – Stellungnahme.
 - 5 Deutsche Forschungsgemeinschaft, Ständige Senatskommission für Grundsatzfragen in der Klinischen Forschung, Arbeitsgruppe „Qualität in der Klinischen Forschung“ (2018): Replizierbarkeit von Ergebnissen in der Medizin und Biomedizin – Stellungnahme.

und Vermeidung (replacement) von Tierversuchen. Maßnahmen im Sinne des 3R-Prinzips und das bei Forschungsprojekten inhärente Ziel, ein Maximum an wissenschaftlicher Aussagekraft zu erzielen, stellen zwei Aspekte dar, die bei der Planung und Beschreibung wissenschaftlicher Projekte nicht losgelöst voneinander betrachtet werden dürfen. Auch stellen die zu erwartende Belastung der Tiere und der angestrebte Erkenntnisgewinn aus dem Forschungsprojekt die Grundlage für eine Güterabwägung dar, ob der Einsatz von Tieren zu Forschungszwecken zu rechtfertigen ist.⁶

Insofern ist es im ureigenen Interesse der Wissenschaft, hohe wissenschaftliche Standards in tierexperimentellen Forschungsprojekten anzustreben und Maßnahmen zur Sicherstellung des Tierschutzes unmittelbar in der Projektplanung zu verankern.

Um diesem Anliegen nachzukommen, beschränkt sich die Handreichung bewusst auf Aspekte des 3R-Prinzips bei der Ausarbeitung, Planung und Durchführung von tierexperimentellen Forschungsprojekten. Konkrete Spannungsfelder werden beispielhaft vorgestellt. Übergeordnete Aspekte, die die Qualitätssicherung in der tierexperimentellen Forschung allgemein betreffen, werden in der Stellungnahme der Arbeitsgruppe „Qualität in der Klinischen Forschung“ thematisiert. Dieser Stellungnahme schließt sich die Senatskommission für tierexperimentelle Forschung an.

6 Deutsche Forschungsgemeinschaft, Ständige Senatskommission für tierexperimentelle Forschung (2016): Tierversuche in der Forschung.

3 Mehr als 3R – 3R-Prinzip und wissenschaftliche Aussagekraft

3.1 Legitimation von tierexperimenteller Forschung durch Güterabwägung

Die Verwendung von Tieren in der Forschung richtet sich nach dem Grundsatz der Verhältnismäßigkeit. Dieser kommt als rechtsstaatliches Prinzip immer dann zur Anwendung, wenn mehrere im Grundgesetz verankerte Grundrechte, Rechtsgüter oder -prinzipien (hier: Allgemeines Persönlichkeitsrecht, Integrität von Leib und Leben, Forschungsfreiheit und Tierschutz) kollidieren. Demnach sind Versuche an Tieren legitim, wenn die zu erwartenden Erkenntnisse bedeutsam sind und den Tieren dabei keine vermeidbaren Schmerzen, Leiden oder Schäden zugefügt werden. Das Instrument zur Überprüfung der Verhältnismäßigkeit eines Tierversuchsvorhabens ist die Güterabwägung im Sinne einer Schaden-Nutzen-Analyse. Dabei wird überprüft, ob ein legitimes Versuchsvorhaben (a) geeignet, (b) erforderlich und (c) angemessen ist, um den angestrebten Erkenntnisgewinn zu erzielen.⁷

Eignung und Erforderlichkeit sind notwendige Voraussetzungen für eine Güterabwägung; hier wird nach der wissenschaftlichen Begründung des Versuchsvorhabens gefragt (Eignung) und nach möglichen Alternativen zur Verwendung von Tieren und belastenden Methoden (Erforderlichkeit). Nur wenn ein Versuchsvorhaben wissenschaftlich sinnvoll und geeignet erscheint und keine Alternativen zu Tierversuchen zur Verfügung stehen, kann eine Güterabwägung im Sinne der Schaden-Nutzen-Analyse erfolgen. Die Güterabwägung konzentriert sich dann auf die Frage der Angemessenheit von Versuchsvorhaben und überprüft, ob der zu erwartende Nutzen des Versuchsvorhabens die Belastungen der Versuchstiere rechtfertigt. Nur dann kann das Versuchsvorhaben genehmigt werden.

7 Wienbracke, M. (2013): Der Verhältnismäßigkeitsgrundsatz. Zeitschrift für das Juristische Studium (ZJS). Nr. 2: 148–155.

3.2 Das 3R-Prinzip als normatives Kriterium für den Tierschutz in der Forschung

Jeder Tierversuch muss nach dem deutschen Tierschutzgesetz (TierSchG) vor Durchführung bei der zuständigen Genehmigungsbehörde beantragt werden (§ 8 TierSchG). Im Zuge des Genehmigungsverfahrens wird ausdrücklich geprüft, ob das 3R-Prinzip (Replace, Reduce, Refine) in seinen unterschiedlichen Aspekten sinnvoll und bestmöglich umgesetzt wird. Hier ist nachzuweisen, dass der angestrebte Erkenntnisgewinn nicht ohne Verwendung empfindungsfähiger Tiere (Replace), mit weniger Tieren (Reduce) oder mit weniger belastenden Methoden (Refine) erzielt werden kann. Neben dem Einsatz von Methoden ohne Versuchstiere (sogenannte Alternativmethoden) ist zu prüfen, ob Tiere mit geringerer Empfindungsfähigkeit eingesetzt werden können („relatives replacement“). Die Minimierung der Tierzahl erfordert eine sorgfältige Versuchsplanung mit Stichprobenberechnung (zum Beispiel Poweranalyse). Zur Minimierung der Belastung der Tiere stehen zahlreiche Verfahren zur Verfügung. So können zum Beispiel eine Verbesserung der Haltungsbedingungen durch Umweltanreicherung, die schrittweise Gewöhnung der Tiere an die Versuchsbedingungen, die Verwendung nicht invasiver Techniken sowie optimierte Anästhesie- und Analgesieverfahren die Gewichtung von Schaden und Nutzen in der Güterabwägung zugunsten eines Tierversuchs verschieben.

3.3 Das 3R-Prinzip im Kontext von Validität und Replizierbarkeit

Eine gewissenhafte Umsetzung der 3R ist eine notwendige, aber nicht hinreichende Voraussetzung für die ethische Vertretbarkeit von Tierversuchen. Unabhängig vom Belastungsgrad der Tiere muss das Studiendesign stringenten wissenschaftlichen Qualitätsanforderungen an Objektivität, Validität und Replizierbarkeit entsprechen. Fundamental für die Verantwortung eines jeden Forschenden ist die Sicherstellung wissenschaftlicher Qualität, um maximale wissenschaftliche Aussagekraft zu erzielen. Dieses Ziel muss ebenso immer zentrales Kriterium des wissenschaftlichen Begutachtungsverfahrens von Forschungsprojekten sein.

Eine positive Bewertung der wissenschaftlichen Aussagekraft (siehe Abschnitt 3.4) eines Versuchsvorhabens und die Anwendung des 3R-Prinzips ist Voraussetzung für die abschließende Prüfung der Angemessenheit im Rahmen der finalen Güterabwägung. Gemäß dem Grundsatz der Verhältnismäßigkeit kann vor allem unter Berücksichtigung des nur in Kollision mit anderen Grundrechten einschränkbareren Grundrechts der Forschungsfreiheit das Kriterium der wissenschaftlichen Aussagekraft dem nach 3R aufgestellten Tierschutz vorgeordnet werden.⁸ Das 3R-Prinzip kann seine Funktion nur bei Vorliegen eines validen Versuchsplans erfüllen und das Reduce-Prinzip darf nicht so missverstanden werden, dass die Tierzahl in einer Studie auf Kosten der Aussagekraft gesenkt wird (siehe Abschnitt 4.2). Ob ein Tierversuch unter Berücksichtigung seiner wissenschaftlichen Qualität und seiner Erforderlichkeit angemessen und damit genehmigungsfähig ist, muss die abschließende Güterabwägung ergeben.

3.4 Aspekte der wissenschaftlichen Aussagekraft von Tierversuchen

Im Gegensatz zum 3R-Prinzip als zentraler Komponente bei der Beurteilung der ethischen Vertretbarkeit der Durchführung von Tierversuchen steht für die Beurteilung der Eignung von Tierversuchen kein ähnlich griffiges Konzept zur Verfügung. Ganz generell lässt aber die Prüfung der wissenschaftlichen Aussagekraft von Versuchsergebnissen eine gute Orientierung zu. Die wesentlichen Kriterien wissenschaftlicher Aussagekraft von tierexperimentellen Forschungsprojekten lassen sich drei unterschiedlichen Aspekten wissenschaftlicher Validität zuordnen, die sich auf die Güte des Tier- beziehungsweise Versuchsmodells, die Güte der postulierten Ursache-Wirkung-Relation und auf den Grad der Generalisierbarkeit beziehen. Diese werden beispielsweise nach Würbel (2017) in Konstruktvalidität, interne Validität und externe Validität untergliedert:⁹

8 Wienbracke, M. (2013): a. a. O.

9 Würbel H. (2017): More than 3Rs: The importance of scientific validity for harm-benefit analysis of animal research. *Lab Animal*, 46(4): 164–166.

Konstruktvalidität bezieht sich auf die Güte des Tier- beziehungsweise Versuchsmodells und der verwendeten Methoden zur Untersuchung der konkreten Fragestellung eines Versuchsvorhabens. Sie bezeichnet die Genauigkeit, mit der eine Versuchsanordnung das misst, was sie zu messen vorgibt.¹⁰ Die Beurteilung der Konstruktvalidität sollte sich deshalb auf empirische Evidenz zur biologischen Übereinstimmung zwischen Eigenschaften der Versuchspopulation und Eigenschaften der Zielpopulation und auf die Aussagekraft der Zielgröße des Versuchs für die damit zu messende Funktion beziehen.^{11,12}

Interne Validität bezeichnet die Güte der postulierten Ursache-Wirkung-Relation. Interne Validität ist gegeben, wenn eine Veränderung in der abhängigen Variable (ein Behandlungseffekt) nachweislich auf Variation in der Versuchsbehandlung zurückgeführt werden kann. Eine Beurteilung der internen Validität muss deshalb neben grundlegenden Aspekten der Versuchsplanung und der Verwendung von Kontrollgruppen auch alle anwendbaren Kriterien guter wissenschaftlicher Praxis zur Vermeidung von Verzerrungen (bias) berücksichtigen (zum Beispiel Randomisierung, Verblindung, Stichprobenberechnung, Definition der primären und sekundären Zielvariablen, Definition von Ein- und Ausschlusskriterien, Definition des statistischen Auswertungsplans etc.).

Externe Validität bezeichnet den Grad der Generalisierbarkeit von Versuchsergebnissen über die spezifischen Bedingungen des konkreten Versuchs hinaus. In der empirischen Forschung steht die Generalisierbarkeit der Befunde im Vordergrund. Allein für die Replizierbarkeit von Ergebnissen ist ein

10 Koob, G. F., Heinrichs, S. C., Britton K. (1998): Animal models of anxiety disorders. The American Psychiatric Press Textbook of Psychopharmacology. (2. Schatzberg AF, Nemeroff CB, editor). Washington DC-London: American Psychiatric Press: 133–144.

11 Willner, P. (1984): The validity of animal models of depression. Psychopharmacology, 83(1): 1–16.

12 Belzung, C., Lemoine, M. (2011): Criteria of validity for animal models of psychiatric disorders: focus on anxiety disorders and depression. Biology of Mood & Anxiety Disorders, 1(1): 9.

Mindestmaß an externer Validität erforderlich, da sich die Versuchsbedingungen zwischen Wiederholungsversuchen (sowohl innerhalb eines Versuchslabors als auch zwischen verschiedenen Versuchslaboren) zwangsläufig unterscheiden. Ausgehend davon sollte die Beurteilung der externen Validität Aspekte des Versuchsplans berücksichtigen, die Rückschlüsse auf die Replizierbarkeit und Generalisierbarkeit (zum Beispiel auf andere Tiermodelle, andere Versuchsbedingungen) zulassen. Dazu gehören unter anderem die Berücksichtigung der Geschlechter, die Aufteilung eines Versuchs in mehrere unabhängige Replikate, die systematische Variation einer oder mehrerer unabhängiger Variablen (zum Beispiel mehrere Mauslinien, Haltungsbedingungen, Versuchsanordnungen oder Experimentatoren) oder die Durchführung von Multi-Labor-Studien.

Mit diesen drei Aspekten der Validität tierexperimenteller Forschungsergebnisse können wesentliche Kriterien zur Beurteilung der wissenschaftlichen Aussagekraft und Replizierbarkeit von Tierversuchen abgedeckt werden. Analog zum 3R-Prinzip, welches eine Beurteilung der Erforderlichkeit von Tierversuchen ermöglicht, kann die Anwendung dieser drei Aspekte der Validität die Beurteilung der Eignung von Tierversuchsvorhaben zur Erzielung des angestrebten Nutzens erleichtern.

4 Spannungsfelder zwischen der wissenschaftlichen Validität und den 3R

Grundsätzlich haben Forschende nicht nur aus ethischer Sicht ein großes Interesse daran, Tierschutzmaßnahmen in ihren Projekten umzusetzen. So können tierschutzrelevante Belastungen (Schmerzen, Leiden, Schäden) der Versuchstiere die Aussagekraft der Ergebnisse beeinträchtigen. Deshalb ist es nicht nur aus Tierschutzgründen, sondern auch mit Blick auf die Qualität der Forschung ein vorrangiges Anliegen, die Belastungen der Tiere soweit wie möglich zu minimieren. Das 3R-Prinzip und wissenschaftliche Aussagekraft gehen in den meisten Fällen Hand in Hand.

In manchen Bereichen der Versuchsplanung bestehen jedoch grundlegende Spannungsfelder zwischen der Umsetzung des 3R-Prinzips und der Maximierung der wissenschaftlichen Aussagekraft. Dies kann sich auch auf die Beurteilung von Versuchsvorhaben im Rahmen des Genehmigungsverfahrens auswirken. Die Anwendung verbindlicher Kriterien wissenschaftlicher Validität (zum Beispiel Konstruktvalidität, interne und externe Validität) helfen, die Auswirkungen von 3R-Maßnahmen auf wesentliche Aspekte der wissenschaftlichen Aussagekraft zu prüfen.

Spannungsfelder bestehen bezüglich aller 3R; bei der Verwendung bestimmter Tierarten, bei der Festlegung der Tierzahl (Stichprobengröße) und bei der Anwendung bestimmter tierexperimenteller Techniken. Um Forschenden bei der Planung, Beschreibung und Begründung ihrer Projekte eine Hilfestellung zu geben, werden einige dieser Spannungsfelder exemplarisch dargestellt und mögliche Lösungsansätze vorgeschlagen.

4.1 Wahl der Tierart

Der Begriff „Replace“ meint nicht nur den Ersatz von Tierversuchen durch Alternativmethoden (in silico, in vitro usw.), sondern umfasst auch „relatives replacement“, das heißt das Ausweichen auf sogenannte „niedere“ Tiere (Wirbellose) oder „niederere“ Wirbel- bzw. Säugetiere (zum Beispiel Mäuse statt

Primaten, Fische statt Säugetiere). Hintergrund für das Verständnis des „relativen replacements“ ist die verbreitete Auffassung, dass sogenannte „höhere“ Wirbeltierarten (zum Beispiel Primaten, aber auch Hunde, Katzen) unter den Versuchseinwirkungen stärker leiden als „niederere“ Arten¹³. Diese Auffassung spiegelt sich auch in der EU-Richtlinie 2010/63/EU zum Schutz der für wissenschaftliche Zwecke verwendeten Tiere wider, die Primaten und andere Tierarten unter einen besonderen Schutz stellt. Die postulierte Hierarchie des Leidens „höherer“ versus „niederer“ Arten unter gegebenen Versuchsbedingungen ist aus wissenschaftlicher Sicht fragwürdig und wird auch aus ethischer Sicht kontrovers beurteilt.^{14,15} Das Tierschutzgesetz geht grundsätzlich bei allen Wirbeltieren (und einigen Wirbellosen) von Empfindungsfähigkeit aus und schließt diese für andere Tierarten nicht aus¹⁶. Für eine ethische Hierarchisierung innerhalb der Wirbeltiere anhand unterschiedlicher Leidensfähigkeit fehlen bisher verbindliche Kriterien. Die Anwendung des „relativen replacement“ kann auch aus wissenschaftlicher Sicht durchaus kritisch sein, wenn die Wahl des Tiermodells einen Einfluss auf die wissenschaftliche Aussagekraft hat. Der Nutzen der Forschung (Erkenntnisgewinn) und damit auch deren ethische Legitimation kann dadurch unter Umständen beeinträchtigt werden.

- ▶ *Neben einer ethischen und juristischen Klärung der postulierten ethischen Hierarchie innerhalb der vom Tierschutzgesetz geschützten Tierarten braucht es diesbezüglich auch eine größere Sensibilisierung für die Problematik. Forschende sollten sich bei der Versuchsplanung bewusst mit der Wahl ihrer Tiermodelle auseinandersetzen und diese im Rahmen der Güterabwägung auch anhand wissenschaftlicher*

13 So schreibt das Tierschutzgesetz (§ 7(1)1) vor, dass „Tierversuche im Hinblick auf [...] die artspezifische Fähigkeit der verwendeten Tiere, unter den Versuchseinwirkungen zu leiden, auf das unerlässliche Maß zu beschränken“ sind.

14 Rippe, K. P. (2003): Tierethik. Bioethik, hrsg. v. M. Düwell & K. Steigleder, Frankfurt a. M.

15 Ach, J. S., Borchers, D. (Hg.): Handbuch Tierethik. Grundlagen – Kontexte – Perspektiven. Stuttgart: J. B. Metzler 2018.

16 Tierschutzgesetz § 8(a)(4).

Kriterien der Validität und der artspezifischen Leidensfähigkeit unter den gegebenen Versuchsbedingungen begründen.

4.2 Festlegung der Tierzahl

Der Begriff „Reduce“ bezieht sich hauptsächlich auf die Anzahl von Tieren, die für einen Versuch verwendet werden (Gesamtstichprobe). Aus wissenschaftlicher Sicht wären große Stichproben wünschenswert, aus Sicht des Tierschutzes hingegen möglichst kleine. Die Berechnung der erforderlichen Stichprobengröße (zum Beispiel mittels Poweranalyse) ist in vielen Fällen ein geeignetes Mittel, um die minimale Tierzahl festzulegen.¹⁷ Da der vermeintliche Fortschritt im Tierschutz in der Forschung primär anhand der jährlichen Versuchstierstatistik gemessen und öffentlich kommuniziert wird, besteht zudem erheblicher politischer Druck auf die Verminderung der Tierzahlen. Aufgrund der oben genannten verbreiteten Ansichten einer ethischen Hierarchisierung innerhalb der Wirbeltiere trifft dies im besonderen Maße auf Versuche an sogenannten „höheren“ Säugetieren zu.

Übermäßiges Minimieren der Stichprobengröße unter dem Titel 3R (Reduce) wird dann problematisch, wenn dies durch subkritische wissenschaftliche Aussagekraft „erkauft“ wird. Dies ist besonders dann der Fall, wenn auf wichtige Vergleichs- oder Kontrollgruppen verzichtet wird oder zu kleine Stichproben verwendet werden. Dabei ist zu bedenken, dass die ethischen „Kosten“ von zu geringen Stichproben höher ausfallen können als die von zu großen, da Tierversuche unterhalb einer bestimmten Stichprobengröße aufgrund mangelnder statistischer Power im ungünstigsten Fall wertlos werden können.

Forschende müssen deshalb die Tierzahlen für ihre Versuche anhand wissenschaftlicher Kriterien (etwa der internen und externen Validität der angestrebten Ergebnisse) begründen können. Es gibt zahlreiche Hilfsmittel zur Versuchsplanung (zum Beispiel der Experimental Design Assistant,

17 Festing, M. F. W. (2018): On determining sample size in experiments involving laboratory animals. *Laboratory Animals*, 52(4): 341–350.

<https://eda.nc3rs.org.uk/>) oder zur Berechnung und Begründung der Stichprobengröße (etwa G*Power, www.gpower.hhu.de/). Bei vielen Versuchen besteht zudem die Möglichkeit eines adaptiven Versuchsdesigns (beispielsweise gruppensequentielles Design, Zwischenanalyse mit Anpassung der Stichprobengröße, adaptive Randomisierung usw.), um die Tierzahl ohne Verlust an wissenschaftlicher Aussagekraft zu minimieren. Bei komplexeren Versuchsansätzen empfiehlt es sich jedoch, professionelle Unterstützung durch Biostatistiker mit Expertise in der Tierversuchsplanung in Anspruch zu nehmen.

- ▶ *Die Angaben zur Anzahl von Versuchstieren bedürfen einer sorgfältigen Abwägung zwischen Bestrebungen zur Reduktion der Tierzahlen und der wissenschaftlichen Aussagekraft und sollten bei der Beschreibung des Projekts erläutert werden.*

4.3 Standardisierung

Tierversuche werden meist im Versuchslabor unter kontrollierten Bedingungen durchgeführt. Kontrollierte Versuchsbedingungen sind in der Grundlagenforschung oft wünschenswert, um allfällige Störvariablen auszuschließen. Dabei werden die Tiere und die Haltungsbedingungen meist rigoros standardisiert, indem genetisch weitgehend identische Tiere (zum Beispiel Inzuchtlinien) unter identischen Bedingungen (bezüglich Käfig, Futter, Management) aufgezogen und gehalten werden. Standardisierung wird zudem oft als Mittel zur Minimierung der Tierzahl (Reduce) empfohlen, da sich Behandlungseffekte bei geringerer Variation der Versuchsergebnisse mit kleineren Stichproben statistisch nachweisen lassen.

Standardisierung kann zwar die Präzision von Versuchsergebnissen verbessern (geringere Variation), übermäßige Standardisierung beeinträchtigt allerdings die externe Validität und damit die Generalisierbarkeit der Versuchsergebnisse.¹⁸ Durch rigorose Standardisierung der Versuchsbedingungen lässt

18 Richter, S. H. et al. (2009): Environmental standardization: cure or cause of poor reproducibility in animal experiments? *Nature Methods*, 6: 257–261.

sich deshalb allenfalls die Tierzahl pro Versuch minimieren, dafür sind mehr unabhängige Versuche – und damit letztlich mehr Tiere – erforderlich, um die Reproduzierbarkeit bei variierenden Bedingungen und damit die Generalisierbarkeit nachzuweisen.¹⁹

Forschende sollten deshalb Versuche möglichst so planen, dass auch Rückschlüsse auf die externe Validität (das heißt auf die Generalisierbarkeit über die Geschlechter sowie genetische und umweltbedingte Variation) möglich sind. Dazu eignet sich die kontrollierte Heterogenisierung der Versuchspopulation mittels mehrfaktorieller Versuchsdesigns, mit denen gezielt die Geschlechter, mehrere Genotypen und/oder mehrere Umwelt- beziehungsweise Versuchsbedingungen berücksichtigt werden können, ohne dass dafür mehr Tiere erforderlich sind.²⁰ Besonders robuste Ergebnisse können mit Multi-Labor-Studien erzielt werden.²¹ Obwohl nicht immer umsetzbar, erscheint dies insbesondere in großen Forschungsverbänden eine sinnvolle Ergänzung der Forschungsmethodik.

- ▶ *Forschende sollten bei der Versuchsplanung externe Validität, das heißt die Replizierbarkeit und Generalisierbarkeit berücksichtigen und das Studiendesign entsprechend begründen.*

4.4 Pilot- und Explorationsversuche

Viele Versuchsvorhaben sind Teil umfangreicher Forschungsprogramme. Dabei werden in einem ersten Schritt oft explorative Studien zur Generierung von Hypothesen durchgeführt, gefolgt von Pilotstudien zur Optimierung des Versuchsplans im Hinblick auf konfirmatorische Studien zur Über-

19 Voelkl, B., Vogt, L., Sena, E. and Würbel, H. (2018): Reproducibility of preclinical animal research improves with heterogeneity of study samples. *PLoS Biology*, 16(2), e2003692.

20 Shaw, R., Festing, M. F. W., Peers, I. and Furlong, L. (2002): Use of Factorial Designs to Optimize Animal Experiments and Reduce Animal Use. *ILAR Journal*, 43(4): 223–232.

21 Voelkl, B., Vogt, L., Sena, E. and Würbel, H. (2018): a. a. O.

prüfung konkreter Hypothesen. Für Explorations- und Pilotstudien sowie in manchen Bereichen der Grundlagenforschung können keine genauen Stichprobenberechnungen durchgeführt werden.²² Die Studien dienen jedoch nicht nur der Generierung vielversprechender Hypothesen und der Optimierung von Versuchsdesigns, sondern tragen auch zur Minimierung der Tierzahl im Rahmen sich anschließender umfangreicher Forschungsprogramme bei. Andernfalls besteht das Risiko, unergiebigere Hypothesen zu testen oder Hypothesen mit unausgereiften Versuchsdesigns und ungeeigneten Methoden zu prüfen, wodurch die wissenschaftliche Aussagekraft erheblich beeinträchtigt werden kann.

- ▶ *Explorations- und Pilotstudien sollten so gestaltet werden, dass sie im Hinblick auf weiterführende Studien beziehungsweise umfassende Forschungsprogramme zu einer Fokussierung und damit zu einer Minimierung der insgesamt benötigten Tierzahl führen.*

4.5 Wiederholungsversuche

Gemäß geltendem Tierschutzgesetz sind Tierversuche nur dann unerlässlich, wenn sie den jeweiligen Stand der wissenschaftlichen Erkenntnisse erweitern und der Erkenntnisgewinn die Belastungen der Tiere überwiegt. Dies ist auch bei der Legitimation von Wiederholungsversuchen zur Überprüfung der Replizierbarkeit von Versuchsergebnissen zu berücksichtigen.

Replizierbarkeit ist jedoch ein Grundpfeiler des wissenschaftlichen Erkenntnisgewinns. Ohne Überprüfung von Forschungsergebnissen besteht ein hohes Risiko unergiebigere Folgeversuche auf der Basis vielversprechender, jedoch nicht replizierbarer Ergebnisse. Dies spricht einerseits für die Verwendung von Versuchsansätzen, die eine Abschätzung der externen Validität und Replizierbarkeit von Versuchsergebnissen ermöglichen (beispielsweise mehrfaktorielle Designs, Multi-Labor-Studien; siehe Abschnitt 4.3). Anderer-

22 Mayer, B., Muche, R. (2013): Formal sample size calculation and its limited validity in animal studies of medical basic research. *Tierärztliche Praxis, Ausg K Kleintiere Heimtiere*, 41(6): 367–374.

seits muss jedoch auch eine unabhängige Überprüfung der Replizierbarkeit bestehender Erkenntnisse möglich sein.

- ▶ *Forschende sollten Wiederholungsversuche (zum Beispiel zur Überprüfung der Replizierbarkeit von Versuchsergebnissen) in Erwägung ziehen und begründen. Unter Umständen können notwendige Wiederholungsversuche als Bestandteil neuer Versuche in Form von Positiv- oder Negativkontrollen eingeplant werden.*

5 Anhang

Reporting Guidelines

ARRIVE Guideline (NC3Rs)

Kilkenny, C. et al. (2010): *Improving bioscience research reporting: The ARRIVE guidelines for reporting animal research*. PLoS Biology, 8(6), e1000412.

<https://www.nc3rs.org.uk/arrive-guidelines>

HARRP Guidelines (ICLAS)

Osborne, N. et al. (2018): *Improving animal reporting standards*. EMBO Reports, 19(5), e46069.

<http://iclas.org/>

PREPARE Guidelines (NORECOPA)

Smith, A. et al. (2017): *PREPARE: Guidelines for planning animal research and testing*. Laboratory Animals, 52(2).

<https://norecopa.no/prepare>

Übersicht über weitere Reporting Guidelines:

https://www.nlm.nih.gov/services/research_report_guide.html

Datenbanken zur Prä-Registrierung von tierexperimentellen Studien

Animal Study Registry (Bundesinstitut für Risikobewertung)

<https://www.animalstudyregistry.org/>

PreClinicalTrials.EU

<https://www.preclinicaltrials.eu/>

Weitere Information und Links

What are the 3Rs? (NC3Rs)

<https://www.nc3rs.org.uk/the-3rs>

Experimental Design Assistant (NC3Rs)

<https://eda.nc3rs.org.uk/>

G*Power (Heinrich-Heine Universität, Düsseldorf)

<http://www.gpower.hhu.de/>

6 Mitglieder der Ständigen Senatskommission für tierexperimentelle Forschung der Deutschen Forschungsgemeinschaft

Professorin Dr. Brigitte Vollmar, Rostock (Vorsitz)

Professor Dr. Peter Dabrock, Erlangen

Professor Dr. Bernd Fleischmann, Bonn

Professorin Dr. Christiane Herden, Gießen

Professor Dr. Nils Hoppe, Hannover

Professor Dr. Lars Lewejohann, Berlin

Professorin Dr. Heidrun Potschka, München

Professor Dr. Stefan Schlatt, Münster

Professor Dr. René H. Tolba, Aachen

Professor Dr. Stefan Treue, Göttingen

Professor Dr. Hanno Würbel, Bern


Kommissionsekretariat

Dr. Cornelia Exner, Marburg

Dr. Valeska Stephan, Rostock

DFG-Geschäftsstelle

Dr. Sonja Ihle, Bonn



Animal Experimentation in Research: The 3Rs Principle and the Validity of Scientific Research

Guidelines of the Permanent Senate Commission on
Animal Protection and Experimentation of the DFG
for the Design and Description of Animal Experimental
Research Projects

Version: July 2019

Contact:

Dr. Sonja Ihle

Life Sciences 1 Division

Phone: 0228/885-2362

E-Mail: sonja.ihle@dfg.de

Table of Contents

1	Summary	3
2	The Relationship between Animal Welfare in Research and Scientific Quality	4
3	More than 3Rs: The 3Rs Principle and Scientific Validity	6
3.1	Legitimising Animal Experimentation by Means of Harm–Benefit Analysis	6
3.2	The 3Rs Principle as a Normative Criterion for Animal Welfare in Research	7
3.3	The 3Rs Principle in the Context of Validity and Replicability	7
3.4	Aspects of Scientific Validity in Animal Experimentation ...	8
4	Areas of Tension between Scientific Validity and the 3Rs ...	11
4.1	Selection of Animal Species	11
4.2	Determination of the Number of Animals	13
4.3	Standardisation	14
4.4	Pilot Studies and Exploratory Experiments.....	15
4.5	Replicate Experiments	16
5	Appendix	17
6	Members of the Permanent Senate Commission on Animal Protection and Experimentation of the DFG	18

1 Summary

The right to freedom of research as enshrined in the Basic Law for the Federal Republic of Germany and the constitutional objective of animal welfare are two important values with high relevance to animal experimentation in research.

The use of animals for scientific purposes can be ethically justified only if an animal experiment is “indispensable” and the expected scientific outcome justifies the harm imposed on the animals in the course of the experiment.

A crucial ethical guideline in animal experimentation is the 3Rs principle (Replace, Reduce, Refine)¹: Animal experiments may be performed only if no other suitable methods are available to investigate the research question and if the number of animals and the harms imposed on them are limited to the unavoidable minimum. The consistent application of the 3Rs principle is very much in the interest of science itself, since impaired welfare of experimental animals can also compromise the validity of research findings. The effort to maximise the scientific validity and replicability of research findings while observing animal welfare must always be the foundation of study design in animal experimentation.

In practice, tensions may arise between measures taken to advance animal welfare in research and the requirements for ensuring scientific validity. Because of the interdependence of these two aspects, policies for implementing the 3Rs principle should not be considered in isolation, but should rather be integrated in the study design and included in descriptions of research projects.

With these guidelines, the DFG Senate Commission on Animal Protection and Experimentation aims to contribute to the debate on quality in biomedical research and to help define the specific requirements for conducting animal experiments. In addition, this publication supports researchers in the design and adequate description of research projects involving animal experimentation.

1 Russell, W. M. S., Burch, R. L. (1959). The principles of humane experimental technique. London: Methuen.

2 The Relationship between Animal Welfare in Research and Scientific Quality

Scientific research is based on the aspiration to achieve results of the highest possible quality and validity. In discussions surrounding the “replication crisis” in research,^{2,3} the DFG in its statements regarding the replicability of research findings has emphasised the need to ensure the quality of research as a fundamental standard in science. It has therefore initiated a discourse on subject-specific reflection concerning quality and replicability as well as on various aspects of quality assurance.^{4,5}

Animal experimentation must ensure the scientific validity of its findings and at the same time the appropriate and responsible care and use of experimental animals. High animal-welfare standards are a prerequisite for high-quality research, and conversely, animal experiments can only be justified if the quality of the research and thus the validity of its results are guaranteed.

Due to the fundamental imperative to protect animals, researchers bear a special moral responsibility when using animals in experiments. The 3Rs principle provides guiding criteria for animal welfare policies in research by calling for the refinement, reduction and replacement of animal experiments. Measures to implement the 3Rs principle and the objective to maximise scientific validity, which is inherent to all scientific research, should not be considered

-
- 2 Ioannidis, J. (2005). Why most published research findings are false. *PLoS Medicine*, 2(8): 696–701.
 - 3 McLeod, M. R. et al. (2014). Increasing value, reducing waste. *The Lancet*, 383(9912): 101–104.
 - 4 Deutsche Forschungsgemeinschaft (2017). *Replicability of Research Results: A Statement by the German Research Foundation*.
 - 5 Deutsche Forschungsgemeinschaft, Permanent Senate Commission on Key Questions in Clinical Research, Quality in Clinical Research Working Group (2018). *Replizierbarkeit von Ergebnissen in der Medizin und Biomedizin: Stellungnahme der Arbeitsgruppe „Qualität in der Klinischen Forschung“ der DFG-Senatskommission für Grundsatzfragen in der Klinischen Forschung*.

separately. Furthermore, the expected harm imposed on the animals and the intended knowledge gain from the research project provide the basis for a harm–benefit analysis to determine whether the use of animals for research purposes can be justified.⁶

It is thus in the very interest of science to strive for high scientific standards in animal experimentation and to incorporate animal welfare policies in the project design.

To support this endeavour, the focus of these guidelines is deliberately limited to those aspects of the 3Rs principle that are relevant when designing, planning and implementing projects that feature animal experimentation. Specific areas of conflict are illustrated by examples. Overarching issues of quality assurance in animal experimentation in general are addressed in the statement by the “Quality in Clinical Research Working Group”, which is endorsed by the Senate Commission on Animal Protection and Experimentation.

6 Deutsche Forschungsgemeinschaft, Permanent Senate Commission on Animal Protection and Experimentation (2016). Animal Experimentation in Research.

3 More than 3Rs: The 3Rs Principle and Scientific Validity

3.1 Legitimising Animal Experimentation by Means of Harm–Benefit Analysis

The use of animals in research is governed by the principle of proportionality. This rule-of-law principle is applied in cases of conflict between different fundamental rights, legal interests or legal principles under the Basic Law (here: general personal rights, integrity of life and limb, freedom of research, animal welfare). Accordingly, experiments on animals are legitimate if the expected benefit of the research outcomes outweighs the harms inflicted on the animals. The tool for evaluating the proportionality of an animal experiment is the harm–benefit analysis. It entails determining whether an experimental project with a legitimate purpose is (a) suitable, (b) necessary and (c) appropriate in order to achieve the intended gain in knowledge.⁷

Suitability and necessity are prerequisites for performing a harm–benefit analysis; they cover the scientific justification of the experiment (suitability) and potential alternatives to the use of animals and to harmful methods (necessity). Only if an experimental project appears to be scientifically sensible and suitable and if no alternatives to animal experiments are available can a harm–benefit analysis be made. Such an analysis examines whether the experiments are appropriate, that is whether the expected benefit of the project justifies the harm imposed on the experimental animals. Only if this is the case can the experimental project be authorised.

7 Wienbracke, M. (2013). Der Verhältnismäßigkeitsgrundsatz. *Zeitschrift für das Juristische Studium (ZJS)*. No. 2: 148–155.

3.2 The 3Rs Principle as a Normative Criterion for Animal Welfare in Research

According to the German Animal Welfare Act (TierSchG), every animal experiment must be submitted for authorisation to the competent authority before it can be carried out (§ 8 TierSchG). The authorisation procedure includes an explicit review of whether the 3Rs principle (Replace, Reduce, Refine) in its various aspects will be implemented in the best possible way. Applicants must demonstrate that the desired gain in knowledge cannot be achieved without the use of sentient animals (Replace), with fewer animals (Reduce) or with less harmful methods (Refine). The possibility of using methods without experimental animals (alternative methods) or of using animals considered to be less sentient (relative replacement) must be taken into consideration. The minimisation of the number of animals requires careful experimental design with sample size calculation (e.g. power analysis). Numerous methods are available to minimise the harm inflicted on animals. For example, improving housing conditions through environmental enrichment, gradual familiarisation of animals with the experimental conditions, use of non-invasive techniques, and optimised anaesthesia and analgesia methods can shift the harm–benefit ratio in favour of an animal experiment.

3.3 The 3Rs Principle in the Context of Validity and Replicability

Careful implementation of the 3Rs is a necessary but not sufficient prerequisite for the ethical justification of animal experiments. Regardless of the severity of procedures imposed on the animals, the study design must meet stringent scientific quality requirements in terms of objectivity, validity, and replicability. Fundamental to every researcher's responsibility is the assurance of scientific quality with the aim of maximising scientific validity. Furthermore, this aim must always be a key criterion in the scientific review of research projects.

A positive evaluation of the scientific validity (see Section 3.4) and the application of the 3Rs principle are prerequisites for determining the appropriateness

of an experimental project in the final harm–benefit analysis. In keeping with the principle of proportionality, the criterion of scientific validity may be given precedence over the implementation of the 3Rs in view of animal welfare, particularly in consideration of the fundamental right of freedom of research, which can only be restricted to the extent that it conflicts with other fundamental rights.⁸ The 3Rs principle can only fulfil its purpose if there is a suitable experimental design, and the Reduce principle must not be misunderstood in such a way that the sample size of a study is lowered at the expense of scientific validity (see Section 4.2). Whether an animal experiment is appropriate when taking into account its scientific quality and necessity, and can thus be authorised must be determined in the final harm–benefit analysis.

3.4 Aspects of Scientific Validity in Animal Experimentation

In contrast to the 3Rs principle as a central component in the evaluation of the ethical justifiability of animal experimentation, no similarly pithy concept exists for assessing the suitability of animal experiments. In general, however, an examination of the scientific validity of experimental findings can provide good guidance. The explanatory value of experimental findings largely depends on three different aspects of scientific validity, namely the quality of the animal model or experimental model, the quality of the postulated cause–effect relation, and the degree of generalisability. Accordingly, Würbel (2017) categorises these into construct validity, internal validity, and external validity:⁹

Construct validity refers to the validity of the animal model or experimental model and of the methods used to investigate the research question of an experimental project. It describes the accuracy with which an experimental setup

8 Wienbracke, M. (2013). Loc. cit.

9 Würbel H. (2017). More than 3Rs: The importance of scientific validity for harm-benefit analysis of animal research. *Lab Animal*, 46(4): 164–166.

measures what it claims to measure.¹⁰ The assessment of construct validity should therefore focus on empirical evidence of the biological congruence between properties of the test population and those of the target population, as well as on the significance of the primary outcome variable of the experiment for the function to be measured.^{11,12}

Internal validity refers to the quality of the postulated cause–effect relation. Internal validity is present if a change in the dependent variable (a treatment effect) can be demonstrably attributed to variation in the experimental treatment. An assessment of internal validity must therefore take into account not only fundamental aspects of experimental design and the use of appropriate control groups, but also all applicable criteria of good research practice in order to avoid risks of bias (e.g. randomisation, blinding, sample calculation, definition of primary and secondary target variables, definition of inclusion and exclusion criteria, definition of the statistical analysis plan, etc.).

External validity refers to the degree to which experimental results can be generalised beyond the specific conditions of a given experiment. In empirical research, the generalisability of findings often takes priority. The replicability of results alone depends on a minimum of external validity, since experimental conditions inevitably differ between replicate experiments (both within and between experimental laboratories). The assessment of external validity should therefore take into account aspects of the experimental design that allow conclusions to be drawn about replicability and generalisability (e.g. to other animal models, to other experimental conditions). This includes, among others, the

-
- 10 Koob, G. F., Heinrichs, S. C., Britton K. (1998). Animal models of anxiety disorders. *The American Psychiatric Press Textbook of Psychopharmacology*. (2. Schatzberg AF, Nemeroff CB, editor). Washington DC – London: American Psychiatric Press: 133–144.
 - 11 Willner, P. (1984). The validity of animal models of depression. *Psychopharmacology*, 83(1): 1–16.
 - 12 Belzung, C., Lemoine, M. (2011). Criteria of validity for animal models of psychiatric disorders: focus on anxiety disorders and depression. *Biology of Mood & Anxiety Disorders*, 1(1): 9.

consideration of both sexes, the division of an experiment into several independent replicates, the systematic variation of one or more independent variables (e.g. different mouse lines, housing conditions, experimental setups or experimenters), and the conduct of multi-laboratory studies.

These three aspects of the validity of findings from animal experimentation cover essential criteria for assessing the scientific validity and replicability of animal experiments. In analogy to the 3Rs principle, which enables evaluation of the necessity of animal experiments, the application of these three aspects of validity can facilitate the evaluation of the suitability of animal experiments to achieve the intended benefit.

4 Areas of Tension between Scientific Validity and the 3Rs

Researchers have a great stake in implementing animal welfare policies in their projects, and not just for ethical considerations. Conditions that compromise the welfare of laboratory animals (pain, suffering, harm) can impair the validity of research findings. Not only for reasons of animal welfare but also with regard to the quality of research, minimising the harms imposed on animals is therefore a cardinal concern of researchers. The 3Rs principle and scientific validity go hand in hand in most cases.

In some areas of experimental design, however, there is profound tension between the implementation of the 3Rs principle and the maximisation of scientific validity. This can also affect project evaluation in authorisation procedures. The application of binding criteria of scientific validity (such as construct validity, internal and external validity) helps to determine the impact of 3Rs policies on key aspects of scientific validity.

Areas of tension exist with regard to all 3Rs; when it comes to the use of certain animal species, the determination of the number of animals (sample size) and the application of certain experimental methods. In order to help researchers to plan, describe and justify their projects, some of these areas of tension will be illustrated and possible solutions suggested below.

4.1 Selection of Animal Species

Replace does not only refer to the replacement of animal experiments by alternative methods (in silico, in vitro, etc.), but also includes relative replacement, i.e. switching to “lower” animals (invertebrates) or “lower” vertebrates or mammals (e.g. mice instead of primates, fish instead of mammals). The background of the concept of relative replacement is the widespread view that “higher” vertebrate species (such as primates, but also dogs, cats) suffer

more from the effects of experimentation than “lower” species.¹³ This view is also reflected in EU Directive 2010/63/EU on the protection of animals used for scientific purposes, which grants special protection to primates and certain other animal species. The postulated hierarchy of suffering of “higher” versus “lower” species under given experimental conditions is questionable from a scientific point of view and is also controversial from an ethical point of view.^{14,15} The Animal Welfare Act is based on the assumption that all vertebrates (and some invertebrates) are sentient and does not rule this out for other species.¹⁶ There exist as yet no binding criteria to establish an ethical hierarchy within vertebrates based of varying degrees of sentience. Relative replacement can also be quite critical from a scientific point of view if the selection of the animal model impacts scientific validity. In some cases, this may diminish the usefulness of the research (knowledge gain) and thereby also its ethical justification.

- ▶ *In addition to an ethical and legal review of the postulated ethical hierarchy within the animal species protected by the Animal Welfare Act, there is also a need for greater problem awareness. When planning experiments, researchers should carefully consider the choice of their animal models and justify them in the harm–benefit analysis on the basis of scientific validity criteria and the species-specific disposition to suffer under the given experimental conditions.*

13 For example, the Animal Welfare Act (§ 7 (1) (1)) stipulates that animal experiments must be limited to the indispensable minimum, taking into account the experimental animals' species-specific ability to suffer from the effects of the experiment.

14 Rippe, K. P. (2003). Tierethik. Bioethik. Ed. M. Düwell & K. Steigleder. Frankfurt a. M.

15 Ach, J. S., Borchers, D. (eds.) (2018). Handbuch Tierethik: Grundlagen – Kontexte – Perspektiven. Stuttgart: J. B. Metzler.

16 Animal Welfare Act § 8 (a) (4).

4.2 Determination of the Number of Animals

Reduce refers mainly to the number of animals used in an experiment (total sample size). From a scientific point of view, larger samples would often be desirable, but from an animal welfare point of view, smaller samples are preferred. A formal sample size calculation (e.g. by power analysis) is in many cases an appropriate way to determine the minimum number of animals.¹⁷ Since putative progress in animal welfare in research is measured and communicated to the public primarily on the basis of annual statistics on animals used in experimentation, there is also considerable political pressure to reduce the number of animals. Due to the abovementioned views assuming an ethical hierarchy among vertebrates, this applies in particular to experiments involving “higher” mammals.

Excessive minimisation of the sample size based on the 3Rs principle (*Reduce*) becomes problematic if it comes at the price of subcritical scientific validity. This is particularly the case when important comparison or control groups are omitted or samples are used that are too small. It should be borne in mind that the ethical “costs” of samples that are too small can be higher than those of samples that are too large, because in the worst case, animal experiments below a certain sample size may become worthless due to a lack of statistical power.

Researchers must therefore be able to justify the number of animals used in their experiments on the basis of scientific criteria (such as the internal and external validity of the research findings). There are numerous tools for experimental design (e.g. the Experimental Design Assistant, <https://eda.nc3rs.org.uk/>) and for calculating and justifying the sample size (e.g. G*Power, <http://www.gpower.hhu.de/>). In many experiments, it is also possible to use adaptive experimental design (e.g. group-sequential design, interim analysis with adjustment of sample size, adaptive randomisation, etc.) in order to minimise the number of animals without diminishing scientific va-

17 Festing, M. F. W. (2018). On determining sample size in experiments involving laboratory animals. *Laboratory Animals*, 52(4): 341–350.

lidity. For more complex experimental approaches, however, it is advisable to seek professional support from biostatisticians with expertise in animal experimental design.

- ▶ *Determining the number of experimental animals requires a careful balance between striving to reduce the number of animals and ensuring scientific validity. These considerations should be explained in the project description.*

4.3 Standardisation

Animal experiments are usually carried out under controlled laboratory conditions. Controlled experimental conditions are often desirable in basic research in order to exclude any confounding variables. Animals and housing conditions are usually rigorously standardised by housing animals that are genetically largely identical (e.g. inbred lines) under identical conditions (in terms of cage, feed, management). Standardisation has also been recommended as a means of minimising the number of animals (Reduce), since treatment effects can be statistically demonstrated with smaller sample sizes when there is less variation in the experimental results.

Although standardisation can improve the precision of experimental findings (less variation), excessive standardisation affects the external validity and thus the generalisability of the results.¹⁸ Therefore, rigorous standardisation of experimental conditions can at best minimise the number of animals used per experiment, but in turn may necessitate a greater number of independent experiments – and thus ultimately more animals – in order to demonstrate reproducibility under varying conditions and thus generalisability.¹⁹

18 Richter, S. H. et al. (2009). Environmental standardization: cure or cause of poor reproducibility in animal experiments? *Nature Methods*, 6: 257–261.

19 Voelkl, B., Vogt, L., Sena, E. and Würbel, H. (2018). Reproducibility of preclinical animal research improves with heterogeneity of study samples. *PLoS Biology*, 16(2), e2003692.

Researchers should therefore design experiments in such a way that they can also draw conclusions about external validity (i.e. generalisability across both sexes as well as genetic and environmental variation). A suitable method for this is controlled heterogenisation of the experimental population by means of multi-factorial experimental designs, whereby both sexes, several genotypes, and/or several environmental or experimental conditions can be taken into account without requiring more animals.²⁰ Particularly robust results can be achieved with multi-laboratory studies.²¹ Although not always feasible, this appears to be a useful addition to research methodology, especially in large research networks.

- ▶ *When planning experiments, researchers should take into account external validity, i.e. replicability and generalisability, and justify the study design accordingly.*

4.4 Pilot Studies and Exploratory Experiments

Many experimental projects are part of large-scale research programmes. Exploratory studies are often carried out as a first step to generate hypotheses, followed by pilot studies to optimise the experimental design with a view towards confirmatory studies to test specific hypotheses. For exploratory and pilot studies, as well as in some areas of basic research, it is not possible to perform accurate sample size calculations.²² However, the studies not only serve to generate promising hypotheses and optimise experimental designs, but also help to minimise the number of animals used in subsequent large-scale research programmes. Otherwise, there is a risk of testing

20 Shaw, R., Festing, M. F. W., Peers, I. and Furlong, L. (2002). Use of Factorial Designs to Optimize Animal Experiments and Reduce Animal Use. *ILAR Journal*, 43(4): 223–232.

21 Voelkl, B., Vogt, L., Sena, E. and Würbel, H. (2018). *Loc cit.*

22 Mayer, B., Muche, R. (2013). Formal sample size calculation and its limited validity in animal studies of medical basic research. *Tierärztliche Praxis, Ausgabe K: Kleintiere – Heimtiere*, 41(6): 367–374.

unproductive hypotheses or testing hypotheses with immature experimental designs and inappropriate methods, which can significantly impair scientific validity.

- ▶ *Exploration and pilot studies should be designed such that, with a view towards follow-up studies or large-scale research programmes, they enable better focus and thus minimisation of the total number of animals required.*

4.5 Replicate Experiments

According to the Animal Welfare Act, animal experiments are indispensable only if they expand the current state of scientific knowledge and if the gain in knowledge outweighs the harm imposed on the animals. This must also be taken into account when justifying replicate experiments in order to assess the replicability of experimental findings.

However, replicability is a cornerstone of scientific evidence. Without verification of research results, there is a high risk of unproductive follow-up experiments based on promising but non-replicable findings. On the one hand, this underscores the importance of experimental approaches that allow an assessment of the external validity and replicability of experimental findings (e.g. multi-factorial designs, multi-laboratory studies; see Section 4.3). On the other hand, researchers must have the possibility to verify the replicability of previous findings independently.

- ▶ *Researchers should consider and justify replicate experiments (e.g. to assess the replicability of experimental findings). Under certain circumstances, necessary replicate experiments may be incorporated as positive or negative controls into the design of new experiments.*

5 Appendix

Reporting Guidelines

ARRIVE Guideline (NC3Rs)

Kilkenny, C. et al. (2010): *Improving bioscience research reporting: The ARRIVE guidelines for reporting animal research*. PLoS Biology, 8(6), e1000412.

<https://www.nc3rs.org.uk/arrive-guidelines>

HARRP Guidelines (ICLAS)

Osborne, N. et al. (2018): *Improving animal reporting standards*. EMBO Reports, 19(5), e46069.

<http://iclas.org/>

PREPARE Guidelines (NORECOPA)

Smith, A. et al. (2017): *PREPARE: Guidelines for planning animal research and testing*. Laboratory Animals, 52(2).

<https://norecopa.no/prepare>

Overview of additional reporting guidelines:

https://www.nlm.nih.gov/services/research_report_guide.html

Databases for preregistration of animal experimental studies

Animal Study Registry (Federal Institute for Risk Assessment)

<https://www.animalstudyregistry.org/>

PreClinicalTrials.EU

<https://www.preclinicaltrials.eu/>

Further information and links

What are the 3Rs? (NC3Rs)

<https://www.nc3rs.org.uk/the-3rs>

Experimental Design Assistant (NC3Rs)

<https://eda.nc3rs.org.uk/>

G*Power (Heinrich-Heine University Düsseldorf)

<http://www.gpower.hhu.de/>

6 Members of the Permanent Senate Commission on Animal Protection and Experimentation of the DFG

Professor Brigitte Vollmar, Rostock (chair)

Professor Peter Dabrock, Erlangen

Professor Bernd Fleischmann, Bonn

Professor Christiane Herden, Gießen

Professor Nils Hoppe, Hannover

Professor Lars Lewejohann, Berlin

Professor Heidrun Potschka, Munich

Professor Stefan Schlatt, Münster

Professor René H. Tolba, Aachen

Professor Stefan Treue, Göttingen

Professor Hanno Würbel, Bern

Commission secretariat

Dr. Cornelia Exner, Marburg

Dr. Valeska Stephan, Rostock

DFG Head Office

Dr. Sonja Ihle, Bonn



Deutsche Forschungsgemeinschaft

German Research Foundation

Kennedyallee 40 · 53175 Bonn

Postal address: 53170 Bonn

Tel: +49 228 885-1

Fax: +49 228 885-2777

postmaster@dfg.de

www.dfg.de/en

