

Erratum

Zettl/Sieb (Hrsg.)

Diagnostik und Therapie neurologischer Erkrankungen (5. Auflage)

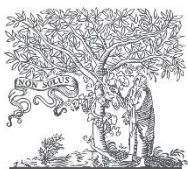
ISBN 978-3-437-21885-9

Liebe Leserin, lieber Leser,

bitte finden Sie hier den Prolog des Buches, verfasst von Professor Ulrich Dirnagl, Berlin.

Wir bedauern das Versäumnis und danken Ihnen für Ihr Verständnis. Wir sind dankbar für jeden Hinweis, der uns hilft, dieses Werk zu verbessern. Bitte richten Sie Ihre Anregungen, Lob und Kritik an folgende E-Mailadresse: kundendienst@elsevier.com

Mit freundlichen Grüßen
Ihre Elsevier GmbH



ELSEVIER

www.elsevier.de

Prolog: Translationale neurologische Forschung – From Bench to Bedside

Hindernisse auf dem Weg vom Labor zum Patientenbett	2
Biologische Komplexität und Krankheitsmodellierung	2
Replikationskrise und Metaforschung	3
Mangelnde interne Validität und Studienqualität	3
Statistikprobleme und unangemessene analytische Flexibilität	4
Geringe externe und Konstrukt-Validität	5
Publikationbias	6
Schlussfolgerung	6

Wichtige Neuigkeiten auf einen Blick

- Eine beträchtliche Anzahl von klinischen Studien scheitert trotz hervorragender Wirksamkeit der Therapie in vorangegangenen und häufig hochrangig publizierten präklinischen Studien. Neben der schieren biologischen Komplexität des Gehirns und seiner Erkrankungen wurden in den letzten Jahren eine Reihe von Ursachen für diese „translationale Blockade“ identifiziert, die Ansatzpunkte zur Überwindung dieser unbefriedigenden Situation bieten.
- Vermutlich ist ein nicht geringer Prozentsatz der Studienergebnisse aus experimentellen Krankheitsmodellen falsch positiv, und damit auch nicht auf den Menschen übertragbar. Hierfür verantwortlich sind mangelnde Kontrolle von Verzerrungen (z. B. Fehlen von Randomisierung und Verblindung oder von vorbestimmten Ein-/Ausschlusskriterien), zu niedrige Fallzahlen sowie statistische Fehler.
- Außerdem haben viele der experimentellen Studien eine zu niedrige externe bzw. Konstruktvalidität. Die im Modell untersuchten Kollektive entsprechen nicht den Eigenschaften, die das Patientenkollektiv charakterisieren (z. B. Geschlecht, Alter, Immunstatus, Komorbiditäten).
- Zudem werden in der experimentellen Medizin negative Studien selten publiziert, die Literatur zeigt einen deutlichen Publikationsbias und eine Überbewertung der positiven Wirkungen von neuen Therapien bei Unterschätzung von Nebenwirkungen und Off-Target-Effekten.
- Es fehlen im experimentellen Bereich kompetente Replikationen von Schlüsselbefunden. Dort, wo dies versucht wurde, konnten die zunächst vielversprechenden Befunde häufig

nicht wiederholt werden oder zeigten deutlich reduzierte Effektstärken.

- Fördergeber, Fachgesellschaften und Journale drängen auf eine Erhöhung der internen und externen Validität der experimentellen Studien, der Qualität der Statistik, auf die Publikation auch von negativen Resultaten sowie die systematische Konfirmation wichtiger Befunde, bevor diese auf Patienten übertragen werden.

- Die Umsetzung dieser Maßnahmen wird zu einem effektiveren Einsatz von Ressourcen führen, nichtinformativ Tierversuche vermeiden helfen sowie Patienten davor schützen, in Studien rekrutiert zu werden, die auf experimenteller Forschung gründet, deren Ergebnisse nicht robust sind.

Hindernisse auf dem Weg vom Labor zum Patientenbett

In unserem Verständnis der Pathophysiologie neurologischer Erkrankungen wurden in den letzten Jahren bemerkenswerte Fortschritte erzielt. Ein Teil dieses Fortschritts gründet auf präklinischer Forschung, die neurologische Erkrankungen *in vivo*, d. h. im Tiermodell, oder *in vitro*, also in Zell- oder Gewebekulturen, modelliert. Die präklinische Forschung orientiert sich dabei häufig an klinischen Fragestellungen und Befunden. Dieses Wechselverhältnis von experimenteller und klinischer Forschung wird auch als „Bench to Bedside Translation“ oder kurz „Translationale Forschung“ bezeichnet. Allerdings scheitert eine beträchtliche Anzahl von klinischen Studien trotz offensichtlicher Wirksamkeit der Therapie in präklinischen Studien. Die Erfolgsquote der Translation ist also beklagenswert niedrig. Selbst von den Medikamenten, die in die Phase I der Humanstudien eintreten, werden später weniger als 10 % als neue Medikamente zugelassen („Ten years on: Measuring the return from pharmaceutical innovation 2019 | Deloitte UK,“ n. d.). Dieser „translationale Block“ existiert für praktisch alle Bereiche der Medizin, ist allerdings besonders ausgeprägt bei manchen neurologischen Erkrankungen. Beispielweise wurden eine Vielzahl von präklinisch hochwirksamen Therapien bei Schlaganfall (Berge et al. 2017) und Alzheimerischer Erkrankung (Egan et al. 2016) beschrieben, von denen sich dann keine einzige bei Patienten als effektiv erwiesen hat.

Es stellt sich daher die Frage, wie es möglich ist, dass präklinische Forschung ganz wesentlich zum grundsätzlichen Verständnis von neurologischen Er-

krankungen beigetragen haben kann, aber bei der Entwicklung von neuen Therapien häufig Probleme auftreten (Dirnagl und Endres 2014; Dirnagl 2016). Die Beantwortung dieser Frage könnte nicht nur helfen, den translationalen Block zu überwinden, sondern auch die Entwicklung lang herbeigesehnter neuer und effektiver Behandlungen für bisher nicht oder nur unzureichend therapierbare Gehirnerkrankungen zu ermöglichen. Wir könnten auch wertvolle Ressourcen einsparen, nichtinformativ Tierversuche reduzieren sowie Patienten davor bewahren, in Studien rekrutiert zu werden, die keinen Nutzen versprechen, aber potenziell riskant sind.

Biologische Komplexität und Krankheitsmodellierung

Vieles liegt bei der Modellierung neurologischen Krankheiten außerhalb der Kontrolle der Forscher – die Physiologie und noch mehr die Pathobiologie des Gehirns ist hoch komplex, oft nicht linear. Häufig spielen auch Faktoren eine wichtige Rolle, die einer Modellierung nicht zugänglich sind, wie etwa Lebensstil und psychische Einflüsse. Experimentelle Modelle können aber selbst unter diesen Umständen durch die Isolation eines spezifischen mechanistischen Signals dennoch zur Aufklärung einzelner Signalwege beitragen oder Hinweise auf die Funktion individueller Gene liefern. Eine Übertragung zum Krankenbett wäre aber wegen der hierbei erhöhten Komplexität dann trotzdem problematisch.

Hinzu kommt, dass die bisherigen Erfolge in der Translation, wie zum Beispiel beim M. Parkinson oder

der Epilepsie, bei denen experimentelle Modelle eine Schlüsselrolle gespielt haben, so etwas wie „niedrig hängende Früchte“ gewesen sind, die wir schon vom Baum der Erkenntnis gepflückt haben. Nun müssen wir uns weiter nach oben strecken – das ist schwieriger und gelingt deshalb seltener. Ein Beispiel ist hier sicher die Alzheimersche Erkrankung, bei der wir nicht einmal gesichert wissen, was sie auslöst und aufgrund welcher Mechanismen Hirnfunktion durch sie verloren geht. Daher ist sie auch schwer einer experimentellen Modellierung zugänglich.

Für eine Reihe von neurologischen Erkrankungen ist dagegen eine experimentelle Modellierung sehr gut möglich, wie zum Beispiel beim ischämischen Schlaganfall. Die akute Blockade eines großen Hirngefäßes ist selbst in Kleinnagern wie Maus oder Ratte relativ einfach herbeizuführen. Es entwickeln sich in der Folge bei diesen Spezies sowohl funktionelle Defizite als auch histopathologische Veränderungen des Gehirns (Infarkt) in analoger Weise zum Menschen. Wieso war aber trotzdem keines der im Tierversuch so hervorragend neuroprotektiv wirkenden Substanzen in den darauf aufbauenden klinischen Studien erfolgreich? Die Antwort auf diese Frage kann auch mithelfen zu verstehen, warum die Replikation der Ergebnisse von wichtigen Studien gerade im präklinischen Forschungsbereich häufig nicht gelingt.

Replikationskrise und Metaforschung

Die Diskussion um die sogenannte „Replikationskrise“ (Baker 2016), die mittlerweile auch Fächer jenseits der Lebenswissenschaften erfasst hat, begann vor etwa 10 Jahren. Die Pharmaindustrie mahnte in mehreren vielbeachteten Publikationen an, dass deren Versuche, publizierte präklinische Forschungsergebnisse „in house“ zu replizieren, häufig scheiterten (Begley und Ellis 2012; Prinz, Schlange und Asadullah 2011). Nun zeigen auch großangelegte, systematische Replikationsstudien wie das „Cancer Reproducibility Project“ („Reproducibility Project: Cancer Biology“, n. d.), dass wohl mehr als 50 % der Schlüsselergebnisse von hochrangig publizierten Studien nicht reproduzierbar sind, und dort, wo dies gelingt, die gemessenen

Effektstärken wesentlich geringer ausfallen als in den Originalpublikationen.

Metaforschung, also Forschung über Forschung (Ioannidis et al. 2015), konnte in den letzten Jahren eine Vielzahl von Faktoren identifizieren, die für diese unbefriedigende Reproduzierbarkeit mit verantwortlich gemacht werden können. Diese spielen vermutlich auch für die Misserfolge in der Translation eine wichtige Rolle. Insofern diese Faktoren fast allesamt von den Forschern kontrolliert werden, besteht die berechnete Hoffnung, dass bei deren Beachtung in den nächsten Jahren eine bessere Reproduzierbarkeit und größere Erfolge in der Übertragung vom Labor in die Klinik erreichbar sein werden.

➤ Tab. unten gibt eine Übersicht über die wichtigsten Faktoren, die die Vertrauenswürdigkeit und Robustheit und damit Reproduzierbarkeit und Translatierbarkeit von experimentellen Befunden negativ beeinflussen können. Keiner davon ist für die präklinische neurologische Forschung spezifisch, aber sie konnten alle in systematischen Reviews und Metaanalysen von neurowissenschaftlichen Veröffentlichungen nachgewiesen werden. Im Folgenden sollen sie kurz beschrieben werden.

Mangelnde interne Validität und Studienqualität

Robustheit und Zuverlässigkeit der Forschung sind durch eine Reihe von Verzerrungen („Bias“) und folglich eine geringe interne Validität gefährdet. Im stark regulierten Bereich der klinischen Forschung ist hohe interne Validität ein zentrales Anliegen bei der Planung, Auswertung und Veröffentlichung von Studien und ein Schlüsselkriterium von Prüfgremien und Aufsichtsbehörden. Überraschenderweise wird in der experimentellen Biomedizin interne Validität viel weniger thematisiert und wird auch nicht von Regulatoren bei der Bewertung der präklinischen Evidenz in Zulassungsverfahren abgefragt oder bewertet (Wieschowski et al. 2018). Dabei wäre eine hohe interne Validität in der präklinischen Forschung relativ einfach sicherzustellen (Macleod et al. 2015).

Selektionsverzerrungen können durch Randomisierung kontrolliert werden. Randomisierung garantiert,

Tab. Wichtige Faktoren, die die Vertrauenswürdigkeit und Robustheit von experimentellen Befunden negativ beeinflussen

Faktor	Beschreibung
Mangelnde interne Validität	Detektions-, Performance-, Konfirmations- und Attrition-Bias führen zu falsch positiven Befunden
Statistische Power/Fallzahl	Zu niedrige Fallzahlen führen zu falsch negativen und falsch positiven Resultaten und einer Überschätzung von tatsächlichen Effektstärken
Mangelnde externe und Konstruktvalidität	Mangelnde Berücksichtigung von Alter, Geschlecht, Komorbidität, Immunsystem, Mikrobiom etc. führen zu eingeschränkter Generalisierbarkeit der Befunde, erniedrigter Replikationswahrscheinlichkeit und fehlender Übertragbarkeit der Ergebnisse auf das intendierte Patientenkollektiv
Statistik- und Analysefehler	p-Hacking, HARKING, verborgene analytische Flexibilität, nichtinformativ Signifikanzniveaus, nicht-korrigierte multiple Vergleiche produzieren falsch positive Resultate
Publikationsbias („Schubladen-effekt“)	Nichtveröffentlichung von negativen oder Null-Resultaten zeichnet ein zu positives Bild von der Wirksamkeit untersuchter Therapien und verhindert robuste Evidenzsynthese in systematischen Reviews

dass die Versuchsgruppen bis auf die experimentelle Manipulation ähnlich sind. Verblindung, also die Geheimhaltung der Versuchsgruppenzuordnung, verhindert den sogenannten Performance Bias. Eine verblindete Bewertung der Ergebnisse verhindert den Detektionsbias. Auch der unkontrollierte Aus- oder Einschluss von Versuchstieren oder experimenteller Ergebnisse, der häufig erst bei der Auswertung erfolgt (sog. Attrition-Bias) kann zu massiver Verzerrung der Ergebnisse führen und allein falsch signifikante Ergebnisse oder eine Überschätzung der Effektstärken erklären (Holman et al. 2016). Er ist durch vorherige Festlegung eindeutiger Ein- und Ausschlusskriterien einfach zu kontrollieren. Das konzeptionelle Gerüst all dieser Maßnahmen wurde bereits vor Jahrzehnten für klinische Studien entwickelt. Leider hat sich trotz einer Reihe von Maßnahmen, wie Auflagen der Fördergeber oder der Journale, in der experimentellen Biomedizin die Situation in den letzten 10 Jahren wenig verbessert. Eine Metaanalyse der gesamten präklinischen Literatur der letzten 10 Jahre (1.381.072 Studien) fand, dass nur etwa 10 % der präklinischen Studien Randomisierung erwähnen, 30 % Verblindung und ein kaum nachweisbarer Anteil eine statistische Powerkalkulation (s. u.) (Menke et al. 2020).

Statistikprobleme und unangemessene analytische Flexibilität

Die Gruppengrößen in der präklinischen Medizin sind traditionell außerordentlich klein. Eine Analyse von mehr als 2.000 experimentellen Schlaganfallstudien, die in den letzten Jahrzehnten durchgeführt wurden, ergab eine mittlere Gruppengröße von 8. Die statistische Power der präklinischen Schlaganfall-Literatur liegt irgendwo zwischen 20 und 45 % (Dirnagl 2006; Schmidt-Pogoda et al. 2020). Sie ermöglicht damit bestenfalls nur eine Vorhersagestärke, die auch durch Werfen einer Münze erreicht würde (50 %). Zum Beispiel bedeutet eine statistische Power von 40 % nicht nur, dass ein Effekt einer vorbestimmten Größe, wenn er tatsächlich vorhanden ist, nur in 40 % der Fälle entdeckt werden kann (d. h. folglich sehr hohe Falsch-Negativ-Rate). Es bedeutet aber auch, dass in den statistisch signifikanten Experimenten die Effektgrößen um mehr als 50 % überschätzt werden (sog. „Winner’s curse“ [Button et al. 2013] = „Der Fluch des Gewinners“). Darüber hinaus, und das ist besonders besorgniserregend, führt eine Power von nur 40 % bei realistischen Prätest-Wahrscheinlichkeiten für die Wirksamkeit der getesteten Substanzen (oder Hypothesen) zu Falsch-Positiv-Raten von mehr als 50 % (Colquhoun 2014). Überwunden werden kann all dies nur durch höhere Fallzahlen und stringenter Signifikanzkriterien. Dies mag auf den ersten Blick als unethisch im Sinne einer Erhöhung der Anzahl der

eingesetzten Versuchstiere erscheinen. Gleichzeitig würde es aber zu einer substanziellen Reduktion von Tierversuchen kommen, denn ein nicht unerheblicher Teil der derzeit durchgeführten Versuche ist aus den oben genannten Gründen nicht informativ, sie müssen daher aus ethischen Gründen unterbleiben (Strech und Dirnagl 2019).

Erwähnt werden muss an dieser Stelle auch, dass viele präklinischen Studien Schwächen im Design und in der Analyse zeigen. Häufig werden eine Vielzahl von Parametern gemessen und dann statistische Signifikanzen ohne Korrektur für diese multiplen Vergleiche durchgeführt (Aarts et al. 2014). Noch problematischer ist das sog. p-Hacking, bei dem nicht präspezifizierte statistische Tests und Kontraste so lange „durchprobiert“ werden, bis einer davon die erwünschte Signifikanz liefert (Colquhoun 2017). Das zumeist verwendete Kriterium für Signifikanz in frequentistischen Verfahren, ein p-Wert kleiner 0,05, wird von den meisten Statistikern zudem als zu wenig strikt und für das Proklamieren einer Entdeckung als ungeeignet erachtet (Benjamin et al. 2018). Auch HARKING („Hypothesizing after the results are known“), also die Bestätigung einer Hypothese, die nach Auswertung der Ergebnisse aufgestellt wurde, ist eine in den experimentellen Biowissenschaften wohl nicht selten geübte Praxis. Dies ist allerdings schwer nachweisbar, da Präregistrierungen von Hypothesen und Studiendetails in der präklinischen Forschung praktisch nicht existieren (Dirnagl 2020).

Geringe externe und Konstrukt-Validität

Bei der translationalen Modellierung in Tiermodellen geht es zum einen um die Reduktion von Komplexität. Allerdings sollten gleichzeitig Untersuchungsbedingungen herbeigeführt werden, die sowohl die Pathophysiologie der Erkrankung beim Menschen möglichst exakt widerspiegeln als auch den Populationsseigenschaften der Erkrankten soweit wie möglich entsprechen. Hierzu zählen z. B. Alter und Geschlecht, aber auch Begleiterkrankungen oder Medikation sollten Berücksichtigung finden. Dies insbesondere dann, wenn eine Studie den Anspruch erhebt, eine

potenziell vielversprechende neue Therapie für den Menschen beschrieben zu haben. Es also im Weiteren um die Entscheidung geht, ob auf die präklinische eine klinische Entwicklungsphase folgen soll. Hier geht es nicht nur um den effizienten Einsatz von Ressourcen, sondern vor allem auch um die Verhinderung von Studien, die nur potenzielle Risiken, aber keinen Nutzen für Studienpatienten haben (Yarborough et al. 2018). Ein Beispiel aus der Schlaganfallforschung kann dies verdeutlichen:

Mit wenigen Ausnahmen wird die Schlaganfallforschung an gesunden, sehr jungen, männlichen Inzucht-Nagerstämmen durchgeführt, die unter spezifisch pathogenfreien (SPF) Bedingungen aufgezogen und mit einer für hohe Wurfgrößen und allgemeine Gesundheit optimierten Diät gefüttert werden. Die äquivalente menschliche Kohorte wären demnach gesunde pubertierende Zwillinge, die in wenige Quadratmeter großen Isolatorzelten mit einer Vitamin und Spurenstoff angereicherten Müsli-Diät aufgezogen werden (Dirnagl 2016). Wir alle wissen aber, dass der Schlaganfall sich meist im höheren Lebensalter ereignet, die Patienten sind beiderlei Geschlechts, auch häufig komorbid, und nehmen chronisch verschiedene Medikamente ein sowie waren im Laufe ihres Lebens zahlreichen Krankheitserregern und Antigenen ausgesetzt. Studien mit Auszucht-Stämmen oder mit komorbiden Tieren, mit hohem Lebensalter etc. haben den starken Einfluss dieser Faktoren auf den Ausgang des Schlaganfalls bei Versuchstieren und auf die Wirksamkeit der Behandlung gezeigt. Generell gilt: Je besser die Schlaganfallmodelle den Schlaganfallpatienten modellieren, desto deutlicher war der Verlust der Wirksamkeit der Behandlungseffekte im Vergleich zu den jungen gesunden Nagern (z. B. Jickling und Sharp 2015; O'Collins et al. 2013). Nicht unerwähnt sollen auch neuere Forschungsergebnisse bleiben, die darauf hindeuten, dass der Status des Immunsystems und damit zusammenhängend auch das Mikrobiom der Versuchstiere eine wichtige Rolle für die Translatierbarkeit und Reproduzierbarkeit von präklinischen Studien spielen (Ivanov et al. 2008; Sadler et al. 2017). Die unter besonders reinen Bedingungen gehaltenen Versuchstiere haben selbst in hohem Alter unreife Immunsysteme (Abolins et al. 2017). Sowohl dem innaten als auch dem adaptiven Immunsystem wird aber eine wichtige Rolle selbst bei nicht primär inflammatorischen ZNS-Erkrankungen

(wie der Multiplen Sklerose) zugeschrieben. Auch bei M. Parkinson, Schlaganfall, M. Alzheimer etc. scheinen inflammatorische Mechanismen bei deren Genese und Verlauf eine wichtige Rolle zu spielen. Es mehren sich Hinweise, dass Phänotypen sowie Therapieeffekte in Krankheitsmodellen stark vom Status des Immunsystems (unreif/reif) beeinflusst werden (Hamilton et al. 2020).

bereits in die Verbesserung der Durchführung der Studie einfließen. Bisher werden Studien ja erst nach Beendigung, wenn ohnehin nichts Grundsätzliches mehr geändert werden kann, begutachtet. Registered Reports garantieren außerdem die Veröffentlichung unabhängig vom (möglicherweise negativen) Ergebnis.

Publikationbias

Selbst in der klinischen Medizin, wo Gesetzgeber, Aufsichtsbehörden und Herausgeber von Fachzeitschriften die Präregistrierung von Therapiestudien vorschreiben, werden vermutlich die Ergebnisse von nicht mehr als der Hälfte aller Studien veröffentlicht (Wieschowski et al. 2019). In der präklinischen Medizin, wo es so gut wie keine Präregistrierung gibt, und Studien oft keinen klar definierten Anfang und kein klar definiertes Ende haben, bleibt unbekannt, welche relevante Evidenz produziert, aber nie berichtet wurde. Ausgehend von dem fast vollständigen Fehlen von Studien in der präklinischen Literatur, die die Null-Hypothese ablehnen (Macleod et al. 2015), können wir jedoch spekulieren, dass es einen überaus starken Publikationsbias in Richtung „wirksame Medikamente“ oder „bestätigte Hypothesen“ gibt. Metaanalysen haben eindeutig die Prävalenz von Publikationsbias und dessen Auswirkungen in allen Feldern der Neurowissenschaften dokumentiert. So ist in der experimentellen Schlaganfall-Literatur der Publikationsbias für mindestens 30 % der veröffentlichten Effektgrößen verantwortlich. Journale könnten hier Abhilfe schaffen, indem sie translationale Studien, deren Resultate den Anspruch begründen, eine für den Menschen relevante Therapie grundgelegt zu haben, nur nach Präregistrierung akzeptieren. Dies würde auch unangemessene analytische Flexibilität (z. B. HARKING, p-Hacking s. o.) verhindern (Dirnagl 2020). Das mittlerweile von einer Vielzahl von Journalen angebotene neue Publikationsformat „Registered Reports“ leistet dies in idealer Weise (Nosek et al. 2018). Hierbei werden Studien nicht nach, sondern bevor sie begonnen wurden, von Experten begutachtet. Dadurch können die Hinweise der Gutachter

Schlussfolgerung

Neurowissenschaftliche Grundlagenforscher und translationale Forscher untersuchen biologische und pathophysiologische Mechanismen neurologischer Erkrankungen. Ihr Ziel ist es oft, neue Therapien, Präventionsmaßnahmen oder Diagnosen zur Verbesserung der menschlichen Gesundheit zu entwickeln. Dabei ist enttäuschend, dass die meisten der im Experiment vielversprechenden Therapien versagen, wenn sie an Patienten getestet werden. Obwohl die Ursachen für dieses Scheitern vielfältig sind und oft auch in der Komplexität der zugrundeliegenden Biologie wurzeln, ist klar geworden, dass große Probleme auch in unzulänglicher Studienmethodik liegen. Die „Translationsblockade“ sowie die sogenannte „Reproduzierbarkeitskrise“ haben eine Diskussion über die Zuverlässigkeit und Reproduzierbarkeit der biomedizinischen Forschung im Allgemeinen ausgelöst, wobei diese in den Neurowissenschaften am weitesten fortgeschritten ist. Es gibt starke Hinweise darauf, dass Schwächen bei der Planung, Durchführung, Analyse und (Nicht-)Veröffentlichung von Forschungsprozessen verbreitet sind. Metaforschung hat gezeigt, dass diese Validitätsbedrohungen zu einer Inflation von Effektgrößen und falsch positiven Ergebnissen führen können und folglich die Reproduzierbarkeit und Vorhersagekraft der Forschung verringern. Gleichzeitig erschwert die zunehmende methodische Komplexität in Verbindung mit der immensen Vermehrung von Forschungsergebnissen die Erstellung und Bewertung von qualitativ hochwertiger Evidenz erheblich. Durch die bevorzugte Veröffentlichung „positiver“ Befunde (Publikationsbias) kommt es zu einer zusätzlichen Überschätzung der Wirksamkeit neuer Therapien in Modellen neurologischer Erkrankungen.

Die biomedizinische Grundlagen- und Translationsforschung hat ein Qualitätsproblem. Dies gilt auch für die Neurowissenschaften. Eine konzertierte und orchestrierte Anstrengung der Wissenschaftler, Institutionen, Journale, Fördergeber, Fachgesellschaften und Regulatoren ist dringend erforderlich, um Ressourcen effizient einzusetzen, unnötige Tierversuche zu verhindern und Patienten nicht unnötigem Risiko auszusetzen. Lesern wissenschaftlicher Veröffentlichungen, in denen auf Basis von Experimenten in Modellen neurologischer Erkrankungen neue, hochwirksame Therapien in Aussicht gestellt werden, sei bis dahin angeraten, die vorgelegte Evidenz besonders sorgfältig zu prüfen.

LITERATUR

- Aarts E, Verhage M, Veenliet JV, Dolan CV, van der Sluis S. A solution to dependency: using multilevel analysis to accommodate nested data. *Nature Neuroscience* 2014; 17(4): 491–496. doi.org/10.1038/nn.3648.
- Abolins S, King EC, Lazarou L, Weldon L, Hughes L, Drescher P, et al. The comparative immunology of wild and laboratory mice, *Mus musculus domesticus*. *Nature Communications* 2017; 8. doi.org/10.1038/ncomms14811.
- Baker M. 1,500 scientists lift the lid on reproducibility. *Nature* 2016; 533(7604): 452–454. doi.org/10.1038/533452a.
- Begley C G, Ellis LM. Drug development: Raise standards for preclinical cancer research. *Nature* 2012; 483(7391): 531–533. doi.org/10.1038/483531a.
- Benjamin D J, Berger J O, Johannesson M, Nosek BA, Wagenmakers E J, Berk R, et al. Redefine statistical significance. *Nature Human Behaviour* 2018; 2(1): 6–10. doi.org/10.1038/s41562-017-0189-z.
- Berge E, Salman R A S, van der Worp H B, Stapf C, Sandercock P, Sprigg N, et al. Increasing value and reducing waste in stroke research. *The Lancet Neurology* 2017; 16(5): 399–408. doi.org/10.1016/S1474-4422(17)30078-9.
- Button KS, Ioannidis J P A, Mokrysz C, Nosek B A, Flint J, Robinson E S J, Munafò MR. Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience* 2013; 14(5): 365–376. doi.org/10.1038/nrn3475.
- Colquhoun D. An investigation of the false discovery rate and the misinterpretation of P values. *Royal Society Open Science* 2014; 1–15. doi.org/10.1098/rsos.140216.
- Colquhoun D. The reproducibility of research and the misinterpretation of p-values. *Royal Society Open Science* 2017; 4(12): 171085. doi.org/10.1098/rsos.171085.
- Dirnagl U, Endres M. Found in Translation: Preclinical Stroke Research Predicts Human Pathophysiology, Clinical Phenotypes, and Therapeutic Outcomes. *Stroke* 2014; 45(5): 1510–1518. doi.org/10.1161/STROKEAHA.113.004075.
- Dirnagl U. Bench to bedside: the quest for quality in experimental stroke research. *J Cereb Blood Flow Metab* 2006; 26(12): 1465–1478. Retrieved from www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16525413.
- Dirnagl U. Thomas Willis Lecture: Is Translational Stroke Research Broken, and if So, How Can We Fix It? *Stroke* 2016; 47(8): 2148–2153. doi.org/10.1161/STROKEAHA.116.013244.
- Dirnagl U. Preregistration of exploratory research: Learning from the golden age of discovery. *PLOS Biology* 2020; 18(3): e3000690. doi.org/10.1371/journal.pbio.3000690.
- Egan K J, Vesterinen H M, Beglopoulos V, Sena E S, Macleod MR. From a mouse : systematic analysis reveals limitations of experiments testing interventions in Alzheimer's disease mouse models. *Evidence-based Preclinical Medicine* 2016; 3, 1: 12–23, e00015. doi.org/10.1002/ebm2.15.
- Hamilton SE, Badovinac VP, Beura LK, Pierson M, Jameson SC., Masopust D, Griffith TS. New Insights into the Immune System Using Dirty Mice. *The Journal of Immunology* 2020; 205(1): 3–11. doi.org/10.4049/jimmunol.2000171.
- Holman C, Piper SK, Grittner U, Diamantaras AA, Kimmelman J, Siegerink B, Dirnagl U. Where Have All the Rodents Gone? The Effects of Attrition in Experimental Research on Cancer and Stroke. *PLOS Biology* 2016; 14(1): e1002331. doi.org/10.1371/journal.pbio.1002331.
- Ioannidis JPA, Fanelli D, Dunne D D, Goodman SN. Meta-research: Evaluation and Improvement of Research Methods and Practices. *PLoS Biology* 2015; 13(10): e1002264. doi.org/10.1371/journal.pbio.1002264.
- Ivanov II, Frutos R de L., Manel N, Yoshinaga K, Rifkin DB, Sartor RB, et al. Specific Microbiota Direct the Differentiation of IL-17-Producing T-Helper Cells in the Mucosa of the Small Intestine. *Cell Host and Microbe* 2008; 4(4): 337–349. doi.org/10.1016/j.chom.2008.09.009.
- Jickling GC, Sharp FR. Improving the translation of animal ischemic stroke studies to humans. *Metabolic Brain Disease* 2015; 30(2): 461–467. doi.org/10.1007/s11011-014-9499-2.
- Macleod MR, Lawson McLean A, Kyriakopoulou A, Serghiou S, de Wilde A, Sherratt N, et al. Risk of Bias in Reports of In Vivo Research: A Focus for Improvement. *PLoS Biology* 2015; 13(10): e1002273. doi.org/10.1371/journal.pbio.1002273.
- Menke J, Roelandse M, Ozyurt B, Martone M, Bandrowski A. The Rigor and Transparency Index Quality Metric for Assessing Biological and Medical Science Methods. *iScience* 2020; 23(11). doi.org/10.1016/j.isci.2020.101698.
- Nosek BA, Ebersole CR, DeHaven AC, Mellor DT. The pre-registration revolution. *Proceedings of the National Academy of Sciences of the United States of America* 2018; 115(11): 2600–2606. doi.org/10.1073/pnas.1708274114.
- O'Collins VE, Donnan GA, MacLeod MR, Howells DW. Hypertension and experimental stroke therapies. *Journal of Cerebral Blood Flow and Metabolism* 2013. SAGE PublicationsSage UK: London, England. doi.org/10.1038/jcbfm.2013.88.
- Prinz F, Schlange T, Asadullah K. Believe it or not: how much can we rely on published data on potential drug targets?

- Nature Reviews. Drug Discovery 2011; 10(9): 712. doi.org/10.1038/nrd3439-c1.
- Reproducibility Project: Cancer Biology. (n. d.). Retrieved February 6, 2020, from cos.io/rpcb/.
- Sadler R, Singh V, Benakis C, Garzetti D, Brea D, Stecher B, et al. Microbiota differences between commercial breeders impacts the post-stroke immune response. *Brain, Behavior, and Immunity* 2017; 66: 23–30. doi.org/10.1016/j.bbi.2017.03.011.
- Schmidt-Pogoda A, Bonberg N, Koecke MHM, Strecker JK, Wellmann J, Bruckmann NM, et al. Why Most Acute Stroke Studies Are Positive in Animals but Not in Patients: A Systematic Comparison of Preclinical, Early Phase, and Phase 3 Clinical Trials of Neuroprotective Agents. *Annals of Neurology* 2020; 87(1): 40–51. doi.org/10.1002/ana.25643.
- Strech D, Dirnagl U. 3Rs missing: animal research without scientific value is unethical. *BMJ Open Science* 2019; 3(1): bmjos-2018-000048. doi.org/10.1136/bmjos-2018-000048.
- Ten years on: Measuring the return from pharmaceutical innovation 2019 | Deloitte UK. (n. d.). Retrieved February 16, 2021, from www2.deloitte.com/uk/en/pages/life-sciences-and-healthcare/articles/measuring-return-from-pharmaceutical-innovation.html.
- Wieschowski S, Chin WWL, Federico C, Sievers S, Kimmelman J, Strech D. Preclinical efficacy studies in investigator brochures: Do they enable risk-benefit assessment? *PLoS Biology* 2018; 16(4): 1–11. doi.org/10.1371/journal.pbio.2004879.
- Wieschowski S, Riedel N, Wollmann K, Kahrs H, Müller-Ohlaun S, Schürmann C, et al. Result dissemination from clinical trials conducted at German university medical centers was delayed and incomplete. *Journal of Clinical Epidemiology* 2019; 115: 37–45. doi.org/10.1016/j.jclinepi.2019.06.002.
- Yarborough M, Bredenoord A, D’Abramo F, Joyce NC, Kimmelman J, Ogbogu U, et al. The bench is closer to the bedside than we think: Uncovering the ethical ties between preclinical researchers in translational neuroscience and patients in clinical trials. *PLoS Biology* 2018; 16(6): e2006343. doi.org/10.1371/journal.pbio.2006343.