



<https://www.sokrates-rationalisten-forum.de/>

## **Das zunehmende Verschwinden der erkenntnistheoretischen Methoden aus der Wissenschaft – Ursache vieler politischer Probleme**

### **Teil III**

#### **Fallstricke und Fehlerquellen bei der Überprüfung von Hypothesen**

*„Wenn Sie für Ihr Experiment Statistiken brauchen, dann sollten Sie lieber ein besseres Experiment machen.“ Ernest Rutherford*

Sind Hypothesen formuliert, geht es an die experimentelle Überprüfung<sup>1</sup> ihres Wahrheitsgehalts, oder präziser formuliert gemäß Karl Popper, um die Erhöhung ihrer Wahrheitsähnlichkeit bzw. Wahrheitsnähe<sup>2</sup>. Hier spielen jedoch weitere Faktoren eine Rolle, z.B. die fachlichen und finanziellen Ressourcen. Leider werden gerade in der Medizin sehr viele Forschungsressourcen und Studienteilnehmer<sup>3</sup> „verschwendet“<sup>4</sup>: Durch unbedeutende Fragestellungen<sup>5</sup> oder an sich oft überflüssige Studien an Me-Too-Präparaten, die zumeist nur eine Scheininnovation darstellen, um Patente zu umgehen<sup>6</sup>. Da diese gut bezahlt werden, sind in Forschungseinrichtungen und Universitäten die Kapazitäten oft blockiert. Darüber hinaus binden zahlreiche, oft redundante Reviews und Metaanalysen ohne wissenschaftlichen Informationsgewinn ebenfalls wichtige Ressourcen<sup>7</sup>.

#### ***Tier- und Humanversuche***

In der Biologie und insbesondere in der Medizin kommt hinzu, dass nicht selten Tiere oder Menschen belastet werden müssen, um verwertbare Ergebnisse zu erhalten<sup>8</sup>, z. B. bei Pharmastudien. Das Ausweichen auf Zellkulturen oder Rechenmodelle ist zwar hilfreich, reicht aber

---

<sup>1</sup> Ein Experiment ist eine methodische Untersuchung zur empirischen Gewinnung von Informationen. Diese Versuche können naturwissenschaftlich sein oder bei medizinischen Fragestellungen Studien an Zellkulturen, Tieren oder Menschen betreffen. Dabei gehören auch prospektiv durchgeführte epidemiologische Studien von zufällig ausgewählten und kontrollierten Gruppen (RCT).

<sup>2</sup> Grundsätzlich können Menschen nie endgültig wissen, ob eine Hypothese der objektiven Wahrheit entspricht, weil wir nicht über einen direkten bzw. unmittelbaren Zugang zur objektiv realen Welt verfügen. Wir können jedoch innerhalb unseres erkenntnistheoretischen Rahmens bessere von schlechteren Hypothesen unterscheiden, indem wir sie durch Experimente und der daraus gewonnenen Datenlage überprüfen.

<sup>3</sup> Ein dauerhaftes Problem der klinischen Forschung ist es, genügend Teilnehmer für eine Studie zu finden.

<sup>4</sup> Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. *Lancet*. 2009 Jul 4;374(9683):86-9. doi: 10.1016/S0140-6736(09)60329-9.

<sup>5</sup> Ioannidis JP, Greenland S, Hlatky MA, et al. Increasing value and reducing waste in research design, conduct, and analysis. *Lancet*. 2014 Jan 11;383(9912):166-75. doi: 10.1016/S0140-6736(13)62227-8.

<sup>6</sup> Aronson JK, Green AR. Me-too pharmaceutical products: History, definitions, examples, and relevance to drug shortages and essential medicines lists. *Br J Clin Pharmacol*. 2020 Nov;86(11):2114-2122. doi: 10.1111/bcp.14327.

<sup>7</sup> Ein Beispiel aus der Kardiologie: Siontis KC, Ioannidis JPA. Replication, Duplication, and Waste in a Quarter Million Systematic Reviews and Meta-Analyses. *Circ Cardiovasc Qual Outcomes*. 2018 Dec;11(12):e005212. doi: 10.1161/CIRCOUTCOMES.

<sup>8</sup> de Vries RB, Wever KE, Avey MT et al. The usefulness of systematic reviews of animal experiments for the design of preclinical and clinical studies. *ILAR J*. 2014;55(3):427-37. doi: 10.1093/ilar/ilu043

selten aus, um eine relevante Hypothese zu überprüfen, wenn es um humane Forschung geht. Die biologischen Regelsysteme in Organismen sind so komplex und zudem häufig nicht gut genug bekannt, um Ergebnisse von Zellexperimenten einfach auf den Menschen übertragen zu können.

Deswegen sind zur Abschätzung von toxikologischen oder pharmakologischen bzw. medizinischen Fragen bei Menschen anfangs Tierversuche und später Studien an Menschen unerlässlich. Hier gibt es ein ausgefeiltes Vorgabensystem, das über die Jahrzehnte hinweg immer weiter verbessert wurde und mittlerweile auch gut funktioniert. An sich müssen jetzt alle Studien an Tieren oder Menschen vorab registriert und auch in jedem Fall publiziert werden, wobei es jedoch immer noch gravierende Erfassungsmängel gibt.

Diese Forderung hat gute Gründe: Es hat sich nämlich gezeigt, dass negative Ergebnisse oft gar nicht publiziert wurden und dadurch ein scheinbarer Vorteil eines Medikamentes „belegt“ wurde, der tatsächlich nicht vorhanden war. Ein besonders eindrucksvolles Beispiel sind Antidepressiva, bei denen oft nur die Studien mit positiven Ergebnissen publiziert wurden<sup>9</sup> <sup>10</sup>. Insbesondere in der nicht-wissenschaftlichen oder sogenannten „alternativen Medizin“ ist dieses unethische Vorgehen<sup>11</sup> häufig zu finden<sup>12</sup>.

### ***Statistische Fallgruben***

Zu diesem Thema gibt es inzwischen viel Literatur, auch allgemeinverständliche, zu der man wenig Vorkenntnisse benötigt<sup>13</sup>. Trotzdem werden weiterhin immer noch grobe Fehler in Publikationen gemacht. Auffällig dabei: Fast immer wird dadurch ein Ergebnis besser dargestellt, als es tatsächlich ist. Besonders beliebt ist der problematische p-Wert, der ohne Angabe des Konfidenzintervalls oft wenig aussagt<sup>14</sup>. Um z. B. bei 5% Irrtumswahrscheinlichkeit ( $p = 0,05$ ) ein scheinbar nicht zufälliges Ergebnis zu bekommen, reichen im Mittel 20 Wiederholungen eines ähnlichen Versuchs aus, um einmal rein zufällig ein positives Ergebnis zu bekommen. Beispielsweise kann man 20-mal eine RCT-Studie mit verschiedenen Homöopathika mit hoher Potenz (starke Verdünnung) durchführen. Obwohl alle ein Placebo darstellen, weil weder im Homöopathikum noch im Placebo ein Wirkstoff enthalten ist, wird im Mittel in einer Studie ein p-Wert von  $>0,05$  herauskommen, aus Zufall. Steht dieses Ergebnis zufällig am Anfang einer Serie, wird die Wirkung scheinbar nochmal verstärkt, da es für viele schon als Beweis gilt.

Manchmal kann auch ein extrem positives Ergebnis zufällig entstehen, was dann statistisch wie ein Lottogewinn zu bewerten ist<sup>15</sup>. Wiederholungen des Versuchs bzw. RCTs (randomisierte kontrollierte Studien) zeigen dann aber sofort die Zufälligkeit bzw. den geringen Effekt<sup>16</sup>. Deshalb sind diese unverzichtbar und stellen einen bedeutenden Fortschritt dar! Systematische Übersichtsarbeiten und Metaanalysen via RCT schränken die Überbewertung einzelner Studien massiv ein, indem diese durch ein Kollektiv ähnlicher Studien zur gleichen Frage ersetzt werden.

Das lässt sich auch – neben der immer wieder geforderten Bestätigung von Ergebnissen durch eine Wiederholung von Studien – als retrospektive Wiederholung verstehen: Man stellt vorhandene

---

<sup>9</sup> Turner EH, Matthews AM, Linardatos E, et al. Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med*. 2008 Jan 17;358(3):252-60. doi: 10.1056/NEJMsa065779.

<sup>10</sup> Ioannidis JP. Effectiveness of antidepressants: an evidence myth constructed from a thousand randomized trials? *Philos Ethics Humanit Med*. 2008 May 27;3:14. doi: 10.1186/1747-5341-3-14.

<sup>11</sup> Antes G, Chalmers I. Under-reporting of clinical trials is unethical. *Lancet*. 2003 Mar 22;361(9362):978-9. doi: 10.1016/S0140-6736(03)12838-3. PMID: 12660049.

<sup>12</sup> Mathie RT, Ramparsad N, Legg LA et al. Randomised, double-blind, placebo-controlled trials of non-individualised homeopathic treatment: systematic review and meta-analysis. *Syst Rev*. 2017 Mar 24;6(1):63. doi: 10.1186/s13643-017-0445-3.

<sup>13</sup> Auswahl: Krämer W: Statistik verstehen - eine Gebrauchsanweisung. Piper 2010

<sup>14</sup> Altman DG. Why we need confidence intervals. *World J Surg*. 2005 May;29(5):554-6. doi: 10.1007/s00268-005-7911-0.

<sup>15</sup> 1987 beobachtete man einen enormen pos. Effekt auf die intravenöse Gabe von Magnesium im Asthmaanfall. Dieses wurde erstrangig publiziert: Okayama H, Aikawa T, Okayama et al. Bronchodilating effect of intravenous magnesium sulfate in bronchial asthma. *JAMA*. 1987 Feb 27;257(8):1076-8.

<sup>16</sup> Mega TA, Gugsu H, Dejenie H, et al. Safety and Effectiveness of Magnesium Sulphate for Severe Acute Asthma Management Among Under-five Children: Systematic Review and Meta-analysis. *J Asthma Allergy*. 2023 Mar 3; 16:241-247. doi: 10.2147/JAA.S390389.

Studien in einem systematischen Ansatz neben die interessierende Studie<sup>17</sup>. Ein solches Vorgehen ändert die Erkenntnis aus einer Studie oft massiv. Allerdings sind die Studien in Design und Durchführung immer wieder so unterschiedlich, dass ihre Ergebnisse sich nicht oder nur teilweise für Metaanalysen eignen. Leider wird die unbedingt nötige Sorgfalt bei der Bewertung der Einzelstudien nicht immer eingehalten<sup>18</sup>.

Eine weitere Effektverzerrung entsteht durch das Weglassen der absoluten Risikoerhöhung oder -verminderung: Beim absoluten Risiko geht die Häufigkeit eines Ereignisses mit der Fallzahl einher. Beim relativen Risiko wird nur der Unterschied von zwei absoluten Risiken angegeben<sup>19</sup>.

Ein Beispiel: Wenn eine Brandschutzmaßnahme das Risiko eines Hausbrandes im Jahr von 1 zu 8.000 (0,0125%) auf 1 zu 10.000 (0,01%) senkt, dann scheint der Risikorückgang um 0,0025 Prozentpunkte als geringer Effekt. Der relative Risikovergleich ergibt allerdings eine Risikoreduktion von 20% ( $1 - 0,01/0,0125$ ), was wie ein großer Effekt wirkt. Gerade bei der Publikation von Pharmastudien (insbesondere in den Anzeigen) wird gerne ausschließlich auf die relative Risikoreduktion zurückgegriffen, um so die Wirksamkeit „aufzuhübschen“<sup>20</sup>.

In epidemiologischen Beobachtungsstudien wird häufig versucht den Einfluss von Confoundern herauszurechnen (adjustieren). Das funktioniert in experimentellen Studien, wenn die Confounder genau bekannt sind. Hierzu gibt es zahlreiche statistische Verfahren. Dieses ist jedoch in Beobachtungsstudien nie der Fall, da immer unbekannte Confounder enthalten sind, die zudem je nach Zusammensetzung der Gruppen unterschiedlich sind. Das wird auch mit moderneren Verfahren wie die Meditationsanalyse<sup>21</sup> nicht besser, denn das Grundproblem bleibt. Die Methoden dienen dort nur der Verfeinerung bzw. Erweiterung der Hypothesen.

### ***Signifikanz und Effektstärke***

Unsere Wachsamkeit darf nie nachlassen, denn für jedes Kontrollsystem werden erfahrungsgemäß immer Schlupflöcher gesucht und auch gefunden. So werden z. B. negative Ergebnisse von Pharmastudien immer wieder in völlig unbekanntem Journalen publiziert, damit sie nicht auffallen; positive Ergebnisse hingegen in hochrangigen Journalen, wodurch ein falsches Übergewicht erzeugt wird.

Ein weiterer Trick: Durch große Fallzahlen wird ein eher dürftiges Ergebnis verschönert und aufgebessert: Bei hohen Fallzahlen steigen nämlich automatisch die Signifikanzwerte an. Selbst viele Wissenschaftler wissen nicht, dass der Signifikanzwert lediglich die Wahrscheinlichkeit für die Zufälligkeit eines Ergebnisses angibt. Über die Wirkungsstärke selbst sagt er jedoch überhaupt nichts aus. Es gibt hier andere Bewertungsfaktoren, wie beispielsweise die MICD (Minimal Important Clinically Difference) in der Medizin oder allgemein die Effektstärke<sup>22</sup>. Dieser Bewertungsfaktor ist aber wenig populär, da viele Medikamente wegen der schwachen Wirksamkeit durchfallen würden<sup>23</sup>. Besonders bei großen Fallzahlen in Studien deutet eine schwache Wirkung auf Untergruppen hin, die nicht oder negativ auf das Medikament ansprechen. Je nach zufälliger Verteilung der Untergruppen können dann in der Summe mal positive oder auch negative Resultate entstehen.

---

<sup>17</sup> Drude NI, Martinez Gamboa L, Danziger M, Dirnagl U, Toelch U. Improving preclinical studies through replications. *Elife*. 2021 Jan 12;10:e62101. doi: 10.7554/eLife.62101.

<sup>18</sup> Ioannidis JP. The Mass Production of Redundant, Misleading, and Conflicted Systematic Reviews and Meta-analyses. *Milbank Q*. 2016 Sep;94(3):485-514. doi: 10.1111/1468-0009.12210.

<sup>19</sup> <https://www.aerzteblatt.de/archiv/81375/Risikokommunikation-Nutzen-und-Risiken-richtig-verstehen> (9/2023)

<sup>20</sup> <https://www.deutsche-apotheker-zeitung.de/daz-az/2016/daz-5-2016/or-rr-nnt> (9/2023)

<sup>21</sup> Rijnhart JJM, Lamp SJ, Valente MJ, MacKinnon DP, Twisk JWR, Heymans MW. Mediation analysis methods used in observational research: a scoping review and recommendations. *BMC Med Res Methodol*. 2021 Oct 25;21(1):226. doi: 10.1186/s12874-021-01426-3.

<sup>22</sup> <https://de.wikipedia.org/wiki/Effektst%C3%A4rke> (9/2023)

<sup>23</sup> Z. B. ist eine Zunahme der Wegstrecke im 6min-Gehtest von 10% oft signifikant, aber der Kranke spürt bei Laufen erst eine Wegstreckenverlängerung von ca. 25%, was die Effektstärke anzeigt.

Ein bekanntes Beispiel ist die Gabe von einem schwachen Diuretikum mit kaliumsparender Wirkung (Spironolacton) bei Herzschwäche. Es zeigte sich in einer großen RCT ein Anstieg der Lebenserwartung<sup>24</sup>. Da die Studie erstrangig publiziert wurde, stieg der Verkauf des Medikaments steil an. Erstaunlicherweise stieg damit auch die Zahl der Todesfälle<sup>25</sup>. In einer Nachuntersuchung zeigte sich, dass in der ersten Studie zufällig viele Patienten mit erniedrigten Kaliumspiegel waren, so dass hier Spironolacton günstig war. Ist jedoch der Kaliumspiegel im oberen Normbereich kann es durch Spironolacton zu einer mitunter lebensbedrohlichen Hyperkaliämie kommen.

Da heute nahezu alle Auswertungen und Analysen mittels Datenverarbeitung geschehen, ist der Weg zum Ergebnis kaum mehr nachvollziehbar, insbesondere bei komplizierten statistischen Verfahren, wie sie in der Regel bei großen Datenmengen verwendet werden. Dadurch gibt es neue Manipulationsmöglichkeiten, bis hin zum Daten Dredging<sup>26</sup> oder p-Hacking<sup>27</sup>.

### **Leitlinien**

Viele wissenschaftliche Ergebnisse in der Medizin sind in Leitlinien oder Empfehlungen ausformuliert. Sie haben eine große praktische Bedeutung, denn sie stellen den diagnostischen und therapeutischen Standard in vielen Bereichen dar. Allerdings müssen auch hier erkenntnistheoretische Maßstäbe eingehalten werden, was nicht immer der Fall ist.

Einer der am häufigsten missachteten Punkte in Leitlinien ist der Unterschied zwischen interner und externer Validität. Mit „externer Validität“ ist gemeint, dass Ergebnisse selbst sorgfältig gemachter RCTs (hier hohe interne Validität) nicht ohne weiteres verallgemeinert werden dürfen. Es muss immer geprüft werden, ob die in den hochwertigen Studien eingeschlossenen Patienten auch der Allgemeinheit in den Praxen entsprechen<sup>28 29</sup>. Patientenauswahl bzw. Ein- und Ausschlusskriterien sind mitunter ein „Steuerungsinstrument“, um Ergebnisse in eine bestimmte Richtung zu verschieben. Nur wenn ein Ergebnis zusätzlich durch dessen pathophysiologischen Hintergrund und andere Studien sowie durch die klinische Erfahrung bestätigt wird, dürften in Leitlinien allgemeine Empfehlungen ausgesprochen werden<sup>30</sup>. Dieser Grundsatz wird leider oft missachtet.

Mitunter gibt es in Leitlinien angebliche Evidenzbewertungen für Behandlungen, ohne dass überhaupt wissenschaftliche Beweise dafür vorliegen. Es steht nur deshalb als Empfehlung in den Leitlinien, weil es geübte Praxis<sup>31</sup> ist, oder weil Infrastrukturen erhalten werden sollen. Das eindrucksvollste Beispiel ist die Intubation und Beatmung bei isoliertem Sauerstoffmangel im Blut (Hypoxämie) infolge einer Lungenentzündung (besonders bei COVID-19-Pneumonie), für die überhaupt keine Evidenz vorhanden ist<sup>32</sup>. In vergleichbaren Kohorten in Deutschland wurde in der Pandemie die Todesrate um das ca. vierfache erhöht, wenn über einen Tubus beatmet wurde. Davon handeln zwei Beiträge im Sokrates-Forum<sup>33</sup>.

Um fehlende oder geringe Evidenz zu kaschieren, werden auch gerne zirkuläre Zitationen<sup>34</sup> verwendet, in denen man auf andere Leitlinien verweist. Nur wenige machen sich dann die Mühe herauszufinden, welche Publikationen die entsprechenden Aussagen tatsächlich begründen. Geht

---

<sup>24</sup> Pitt B, Zannad F, Remme WJ et al. The effect of spironolactone on morbidity and mortality in patients with severe heart failure. Randomized Aldactone Evaluation Study Investigators. N Engl J Med. 1999 Sep 2;341(10):709-17. doi: 10.1056/NEJM199909023411001.

<sup>25</sup> Juurlink DN, Mamdani MM, Lee DS. Rates of hyperkalemia after publication of the Randomized Aldactone Evaluation Study. N Engl J Med. 2004 Aug 5;351(6):543-51. doi: 10.1056/NEJMoa040135

<sup>26</sup> Erasmus A, Holman B, Ioannidis JPA. Data-dredging bias. BMJ Evid Based Med. 2022 Aug;27(4):209-211. doi: 10.1136/bmjebm-2020-111584.

<sup>27</sup> [p-Hacking – Wikipedia](#) (9/2023)

<sup>28</sup> Ioannidis JPA. Science with or without statistics: Discover-generalize-replicate? Discover-replicate-generalize? Behav Brain Sci. 2022 Feb 10;45:e23. doi: 10.1017/S0140525X21000054.

<sup>29</sup> Yarkoni T. The generalizability crisis. Behav Brain Sci. 2020 Dec 21;45:e1. doi: 10.1017/S0140525X20001685.

<sup>30</sup> [Auswahl und Kritische Bewertung der Evidenz | Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e. V. \(awmf.org\)](#) (9/2023)

<sup>31</sup> Im Englischen GCP = Good Clinical Practice

<sup>32</sup> Köhler D, Voshaar T, Stais P et al. Hypoxic, anemic and cardiac hypoxemia: When does tissue hypoxia begin? Dtsch Med Wochenschr. 2023 Apr;148(8):475-482. English, German. doi: 10.1055/a-2007-5450.

<sup>33</sup> [Sokrates - Forum/ Archiv \(sokrates-rationalisten-forum.de\)](#) (Intensivmedizin); [Replik Presserklärung DGP](#) (9/2023))

<sup>34</sup> <https://de.wikipedia.org/wiki/Zirkelbezug> (9/2023)

man der Sache nach, merkt man bisweilen, dass es diese Arbeiten schlicht und einfach nicht gibt! Ein Beispiel dafür sind die Sepsisleitlinien<sup>35</sup>.

Besonders schlimm ist diese Unsitte in der Ernährungsmedizin<sup>36</sup>. Hier werden in Leitlinien sehr oft aus Assoziationen scheinbare kausale Zusammenhänge gefolgert und Ernährungsempfehlungen erstellt und zirkulär zitiert. Daraus sind ebenfalls viele Rechtsverordnungen entstanden sowie jede Menge sinnfreier Alarmismus in den Medien.

So ging vor kurzem durch die deutschen und internationalen Medien, dass eine obst- und gemüsereiche Ernährung sich bei Kindern positiv auf ihr seelisches Wohlbefinden auswirkt. Die Fachgesellschaft der Kinder und Jugendärzte hat aus dieser Korrelation eine Kausalität gemacht<sup>37</sup>. Hintergrund ist eine Studie aus Großbritannien, die diese Assoziation gefunden hat, was auch in der Originalpublikation richtigerweise so geschrieben wurde<sup>38</sup>. Es ist sehr wahrscheinlich, dass Kinder, die mehr Obst und Gemüse essen, ein spezielles Elternhaus bzw. soziales Umfeld haben. Ohne weitere Interventionsstudien oder andere experimentelle Untersuchungen ist diese Hypothese zu einem naheliegenden Verzerrungseffekt nicht zu beantworten.

### **Reproduzierbarkeit**

Auch bei experimentellen Studien gibt es bekannte Fallstricke. Zum einen können ohne böse Absicht einfach Fehler in der Methode vorkommen. Es geschieht leider immer wieder mal, dass das Studiendesign so lange abgeändert wird, bis das gewünschte Ergebnis herauskommt<sup>39</sup>. Auch sind Studien nicht immer reproduzierbar. Das kann daran liegen, dass sich die Bedingungen geändert haben. Öfter ist es jedoch fehlende Sorgfalt oder sogar die Manipulation der Methode oder der Daten<sup>40</sup>. Man geht bei psychologischen Untersuchungen mittlerweile davon aus, dass mehr als 50% der Ergebnisse nicht reproduzierbar sind<sup>41</sup>. Nachanalysen bei zellbiologischen Arbeiten zeigen eine Rate, die noch höher liegen soll<sup>42</sup>. Dazu kommt: Überprüfungen von Studienergebnissen werden eher selten unternommen, da sie wenig Reputation bringen. Ein Lösungsansatz sind systematische Literaturrecherchen<sup>43</sup> sowie Vorschläge zur besseren Standardisierung von Experimenten<sup>44</sup>.

Man sollte also immer kritisch bleiben und zuerst auf die wissenschaftliche Plausibilität achten. Bei besonders spektakulären Ergebnissen ist es klug, eine Überprüfung abzuwarten, falls eine Literaturrecherche kein relevantes Ergebnis liefert. In den letzten Jahren haben sich Wissenschaftsskandale immer wieder daraus entwickelt, dass die Ergebnisse nicht reproduzierbar waren. Erst dadurch fielen Fehler oder Fälschungen auf<sup>45</sup>.

---

<sup>35</sup> Köhler D. Supportive Sepsistherapie: Benutzen wir die falsche Zielgröße? Dtsch Med Wochenschr. 2015 Mar;140(6):439-42. German. doi: 10.1055/s-0041-101010.

<sup>36</sup> Ioannidis JPA. The Challenge of Reforming Nutritional Epidemiologic Research. JAMA. 2018 Sep 11;320(10):969-970. doi: 10.1001/jama.2018.11025.

<sup>37</sup> <https://www.kinderaerzte-im-netz.de/news-archiv/meldung/article/gesunde-ernaehrung-reich-an-obst-und-gemuese-wirkt-sich-auch-auf-seelisches-wohlbefinden-aus/> (10/2023)

<sup>38</sup> Hayhoe R, Rechel B, Clark AB et al. Cross-sectional associations of schoolchildren's fruit and vegetable consumption, and meal choices, with their mental well-being: a cross-sectional study. BMJ Nutr Prev Health. 2021 Sep 27;4(2):447-462. doi: 10.1136/bmjnp-2020-000205

<sup>39</sup> Auswahl: <https://www.psychologie-aktuell.com/news/aktuelle-news-psychologie/news-lesen/korruption-in-der-wissenschaft-eine-ueberwiegend-erfolgreiche-strategie.html> (9/2023); <https://www.welt.de/wissenschaft/article134669083/So-dreist-wird-in-der-Wissenschaft-betrogen.html> (9/2023)

<sup>40</sup> Dirnagl U. Rethinking research reproducibility. EMBO J. 2019 Jan 15;38(2):e101117. doi: 10.15252/embj.2018101117.

<sup>41</sup> Open Science Collaboration. PSYCHOLOGY. Estimating the reproducibility of psychological science. Science. 2015 Aug 28;349(6251):aac4716. doi: 10.1126/science.aac4716.

<sup>42</sup> Yamada KM, Hall A. Reproducibility and cell biology. J Cell Biol. 2015 Apr 27;209(2):191-3. doi:10.1083/jcb.201503036.

<sup>43</sup> [https://www.awmf.org/fileadmin/user\\_upload/dateien/downloads\\_regelwerk/20201214\\_Manual\\_Recherche\\_Evidenzsynthesen\\_Leitlinien\\_V2.1.pdf](https://www.awmf.org/fileadmin/user_upload/dateien/downloads_regelwerk/20201214_Manual_Recherche_Evidenzsynthesen_Leitlinien_V2.1.pdf) (9/2023)

<sup>44</sup> Wilson E, Ramage FJ, Wever KE, et al. Designing, conducting, and reporting reproducible animal experiments. J Endocrinol. 2023 Jun 19;258(1):e220330. doi: 10.1530/JOE-22-0330.

<sup>45</sup> [https://de.wikipedia.org/wiki/Betrug\\_und\\_F%C3%A4lschung\\_in\\_der\\_Wissenschaft](https://de.wikipedia.org/wiki/Betrug_und_F%C3%A4lschung_in_der_Wissenschaft) (9/2023)

## **Resümee**

Betrachtet man die erkenntnistheoretischen Fehler in der Wissenschaft, so liegt der überwiegende Teil im Bereich der Hypothesengenerierung bzw. der daraus gezogenen Schlüsse. Auf Basis bloßer Hypothesen, meist Beobachtungsstudien, werden immer wieder ohne belastbare Argumente Ergebnisse bzw. Kausalitäten vorgegaukelt, die nicht belegt sind. Bei der Hypothesenerstellung ist zudem oft eine selektive Auswahl des bereits vorhandenen Wissens erkennbar.

Durch die heute leichtere Erhebung von vielen, aber oft sehr unspezifischen Daten, gibt es die Neigung, gar keine experimentelle Hypothesenüberprüfung mehr durchzuführen. Damit hängt auch eine zunehmende Unkenntnis über die biomedizinischen Grundlagen zusammen, die für erfolgreiche experimentelle Ansätze unerlässlich ist.

Leider hat dieser wissenschaftliche Mangel zu vielen wissenschaftlich nicht begründeten nationalen und internationalen Rechtsverordnungen geführt, auf deren Basis dann große Summen von Geld und Forschungsressourcen verschleudert wurden. Wir sollten und können das besser machen!

Fassung vom 25. Oktober 2023

### *Autoren*

*Prof. Dr. med. Dieter Köhler (ehem. Direktor, Krankenhaus Kloster Grafschaft, Schmallingenberg)*

*Prof. Dr. rer. nat. Gerd Antes (Mathematiker und Medizinstatistiker, Universität Freiburg)*

*Priv. Doz. Dr. Andreas Edmüller (Philosophie, LMU München)*

*Dr. phil. Andreas F. Rothenberger, Fürstfeldbruck*

*Prof. em. Dr. med. Dr. h.c. Peter Nawroth, em. Direktor Innere Medizin I und Klinische Chemie, Univ. Heidelberg*

*Dr. med. Thomas Voshaar (ehem. Chefarzt, Lungen- und Thoraxzentrum Moers; Vorsitzender des Verbandes Pneumologischer Kliniken e.V.)*

*Dr. phil. nat. Gerhard Scheuch (Physiker mit Schwerpunkt Aerosolmedizin)*

*Prof. Dr. med. Dominic Dellweg (Direktor der Klinik für Innere Medizin, Pneumologie und Gastroenterologie, Pius-Hospital Oldenburg)*

*Dr. med. Patrick Stais, LL.M., MHBA (Pneumologe, Lungen- und Thoraxzentrum Moers)*

*Dr. med. Peter Haidl (ehem. Direktor, Krankenhaus Kloster Grafschaft, Schmallingenberg)*

*Dr. med. Thomas Hausen (Hausarzt im Ruhestand)*