Bevor wir in die Details gehen, ein paar Erklärungen für diejenigen Leser, die nicht mit der Terminologie und dem Kontext vertraut sind:

Bias ist ein Fachbegriff dafür, dass Studienergebnisse verzerrt sind. Klassischerweise geht man in der Methodologie davon aus, dass alle möglichen Variablen Ergebnisse verzerren können.

Wenn etwa in einer Gruppe mehr Raucher, mehr Trinker, intelligentere oder ärmere Leute sind, könnte das das Ergebnis beeinflussen und wirksame Interventionen als unwirksam oder unwirksame Interventionen als wirksam erscheinen lassen. Deswegen werden die Studiengruppen durch Zufallszuteilung erzeugt, damit alle diese Variablen möglichst balanciert sind.

Eine andere typische Form von Bias ist mangelnde Verblindung: wenn etwa Patienten oder klinische Beurteiler die Gruppenzuteilung kennen, dann ist die Chance gegeben, dass die Bewertung nicht unverzerrt ist.

Oder wenn der Studienleiter weiss, dass der nächste Patient in der Kontrollgruppe landen wird, dann könnte es sein, dass er heimlich oder unbewusst dafür sorgt, dass der Patient etwas aufgeschoben wird, bis er in die „richtige" Gruppe kommt, etwa, wenn ihm dieser Patient besonders sympathisch ist. Dieser sog. „allocation bias" wird normalerweise dadurch reduziert, dass Computerprogramme diese Zuteilung vornehmen.

Im neuesten Cochrane Review „Industry sponsorship and research outcome" [1] wird allerdings eine Form von Bias beschrieben, dessen Zustandekommen bei genauerem Nachdenken den Glauben ins Wanken bringt, dass Wissenschaft nichts anderes tue als die Wirklichkeit zu erforschen wie sie ist:

Nennen wir diese neue Form von Bias „Industry Bias": Studien, die von der Industrie finanziert werden, haben nämlich laut dieser Review öfter signifikant bessere Ergebnisse, berichten über weniger Nebenwirkungen, und Präparate des Sponsors sind fast 6 mal so wirksam wie Präparate der Konkurrenz, mit der sie verglichen werden.

Keine der oben angesprochenen klassischen Arten der Ergebnis-Verzerrung spielt in dieser neuen Form von Bias eine Rolle. Die meisten Studien, die in diese Meta-Analyse eingingen waren ihrerseits Meta-Analysen von zum Teil Hunderten von einzelnen sog. randomisierten Studien, also von Studien, die ihre Gruppen durch Zufallszuteilung gebildet hatten.

Die Cochrane-Meta-Analyse wurde von der Cochrane-Collaboration durchgeführt, einer Gruppe von Wissenschaftlern, die ohne weitere Interessenbindung möglichst unbeeinflusst und vollständig die Literatur zusammenfasst. Die Cochrane-Reviews gelten als die sorgfältigsten, weil ein klar definiertes Prozedere vorhanden ist. Der Review wird erst

beantragt und ein Protokoll muß vorgelegt werden, wie die Autoren vorgehen wollen. Es wird begutachtet und erst dann können die Autoren weitermachen. Recherchen müssen vollständig sein und auch die Zusammenfassung folgt einem erprobten System.

Wenn es in der klinischen Forschung also zuverlässige Ergebnisse gibt, dann hier, bei den Reviews der Cochrane Collaboration. Die Reviews sind bekannt dafür, dass sie konservativ sind, also Ergebnisse eher unter- als überschätzen, weil oft sehr restriktive Einschlußkriterien gelten.

Was man wissen muss: hier wurde eine Meta-Meta-Studie von anderen, insgesamt 48 Meta-Studien oder Meta-Analysen durchgeführt. Die Datenbasis sind also einzelne Meta-Analysen, die jede für sich oft mehrere hundert Studien zusammenfassten, insgesamt 9.207 Studien über Arzneimittel und medizinische Geräte, die überwiegende Mehrzahl davon randomisierte Studien (es sind auch einige Beobachtungsstudien dabei, weil nur so Nebenwirkungen zuverlässig erfasst werden können).

Die Autoren stellten eine einfache Frage: gibt es Hinweise darauf, dass Studien, die von Firmen bezahlt werden, häufiger positive Ergebnisse und weniger Nebenwirkungen berichten als solche, die z.B. von der öffentlichen Hand finanziert werden? Das ist deswegen wichtig, weil die größere Zahl von wissenschaftlichen Studien mittlerweile von der Industrie bezahlt wird. Anders gesprochen: das medizinisch-klinische Wissen, das wir haben, wurde in der Mehrzahl aus finanziellen Mitteln von Firmen bezahlt, die dieses Wissen ihrerseits wieder nützen können, um damit Geld zu verdienen.

Dagegen ist überhaupt nichts einzuwenden, wenn, wie wir annehmen, die wissenschaftliche Methodik objektiv ist und genau dann, wenn die geltenden methodischen Kriterien eingehalten werden – Randomisiertung, Verblindung, Verblindung der Zuteilung, etc. – und daher auch kein Unterschied zwischen den Ergebnissen von Studien besteht, die von der Industrie oder der öffentlichen Hand finanziert werden.

Und genau hier zeigt diese Analyse: dies ist nicht der Fall. Studien, die von der Industrie gesponsort sind, haben in 24% der Fälle bessere Ergebnisse als staatlich finanzierte. Industriegesponsorte Studien berichten in 87% der Fälle weniger Nebenwirkungen und kommen in 31% der Fälle zu besseren Schlussfolgerungen. Wenn in solchen Studien das Produkt einer Firma gegen ein Vergleichsprodukt getestet wurde, dann hatte der Vergleich bei Finanzierung durch die Industrie eine fast 6fach höhere Erfolgsrate als bei staatlicher Finanzierung.

Man wird nun sagen: das ist doch klar. Aber bitteschön, einmal Nachdenken. Denn es ist alles andere als klar. Alle Studien werden nach dem gleichen methodischen Standard gemacht. Man könnte sagen, die staatlichen Studien sind halt methodisch nicht so gut. Das

ist eher unwahrscheinlich, weil solche Studien in der Regel durchgeführt werden, wenn ein Produkt auf dem Markt ist. Also müssen sie versuchen, methodisch diejenigen Kritikpunkte aufzugreifen, die geäußert wurden und z.B. noch grössere statistische Mächtigkeit aufweisen. Sie haben also eher eine größere Chance, Effekte zu demonstrieren, wenn sie da sind.

Und die Analyse zeigt: die unterschiedlichen Ergebnisse lassen sich nicht durch methodische Artefakte erklären, denn methodisch gesehen waren sogar die industriegesponsorten etwas besser und statistisch besteht kaum ein Unterschied zwischen den Studientypen. An methodischen Unterschieden kann es also nicht liegen.

Es bleiben aus meiner Sicht zwei Erklärungen übrig: entweder liegt ein ziemlich großer Publikationsbias vor, d.h. die Industrie hält negativ ausgegangene Studien systematisch und in großem Ausmaß unter Verschluß. Im Falle der Antidepressiva wurde ja belegt, dass ein Drittel aller Ergebnisse nicht publiziert wurden [2]. Es könnte durchaus sein, dass das überall so ist und dass Ioannidis mit seiner Einschätzung recht hat, dass die meisten Forschungsergebnisse falsch sind [3], eben weil die negativen Ergebnisse unterschlagen werden. Bei staatlich geförderten Studien ist das eher nicht der Fall, weil hier Forscher und Auftraggeber ein Interesse haben ihre Daten zu publizieren.

Die zweite Erklärung wäre abenteuerlich: sie würde implizieren, dass die Intention, der Wunsch des Forschers bzw. hier des Auftraggebers dazu führt, dass das Ergebnis in der gewünschten Richtung ausfällt, trotz aller methodischen Maßnahmen zur Absicherung gegen solche Effekte. Und weil solche klassischen Versuchsleitereffekte durch methodische Sicherungen ausgeschlossen werden, müssten es nicht-klassische, also vielleicht sogar parapsychologische Effekte sein.

Betrachten wir zum Schluß diese zwei Optionen kurz:

Publikationsbias wäre die natürliche, aber auch extrem verstörende Antwort als Erklärung. Denn es würde bedeuten, dass zwischen 25% und 30% aller Studien, also 200 bis 300 Studien aus dem hier interessierenden Universum von Studien, unpubliziert geblieben sind. Und nicht vergessen: jede dieser Studien kostet schätzungsweise ein bis mehrere Millionen. Die Konsequenz daraus wäre, dass man eigentlich der wissenschaftlichen Literatur nur in Grenzen trauen kann und praktisch immer eine Überschätzung vornimmt, die beträchtlich ist.

Nimmt man hinzu, dass es durch die Gier der Presse sich immer auf erste spektakuläre Ergebnisse zu stürzen, nicht aber Korrekturen zu publizieren auch einen öffentlichen Wahrnehmungsbias gibt [4], dann kann man davon ausgehen, dass man eigentlich kaum irgend einer öffentlich proklamierten Information über „medizinische Fortschritte“ trauen

sollte, bis die Information nicht durch weiterführende Replikationen erhärtet ist.

Nimmt man ausserdem hinzu, dass in industriegesponsorten Studien das Nebenwirkungspotenzial beinahe um 90% geringer ausfällt verglichen mit anderen Studien und dass solche Daten in der Regel aus sehr großen Beobachtungsstudien kommen (weil ja seltene Nebenwirkungen nur berichtet werden können, wenn man tausende von Behandlungen dokumentiert), dann liegt hier Datenfälschung in hohem Ausmaß vor.

Die andere Option würde heißen: die Grundannahmen des experimentellen Modells sind falsch, daß man nämlich den Experimentator – in diesem Falle den Sponsor – und seine Intention durch methodische Maßnahmen (Verblindung, Randomisierung, Verdeckung der Zuweisung, Verblindung der Ergebniserhebung) ausschalten kann. Dann hätten wir es mit direktem Einfluß von Intention oder Bewußtsein auf materielle Systeme zu tun.

Keine dieser Optionen ist komfortabel und man kann sich gewisser Weise aussuchen, ob man seinen Glauben an die Solidität wissenschaftlich erzeugter Daten lieber am Felsen der Scylla zerschellen sieht, dem enormen Publikationsbias, oder ihn lieber im Strudel der Carybdis versinken lassen will, der Unmöglichkeit, die Intention des Experimentators aus dem Ergebnis eines Experiments draußen zu behalten.

Den Publikationsbias kann man in den Griff bekommen, indem man nur noch registrierte Studien zulässt und überprüft, was mit den registrierten aber nicht publizierten passiert ist. Ich würde vermuten, dadurch wird ein Teil des Effektes erklärt. Ob damit aber auch der nicht-klassische Versuchsleitereffekt als Erklärung unnötig wird, für den wir und andere den einen oder anderen Hinweis gefunden haben [5-6]?

Vielleicht werden schon die nachfolgenden Generationen über unsere Naivität lachen, mit der wir geglaubt haben, man könne Systeme willkürlich auseinanderreißen und dennoch gültiges Wissen erlangen? Vielleicht sollten wir anfangen, über die Grundfesten unseres Weltbildes nachzudenken?

Quellen & Literatur

1. Lundh, A., Sismondo, S., Lexchin, J., Busuioc, O. A., ., & Bero, L. (2012). Industry sponsorship and research outcome. Cochrane Database of Systematic Reviews(12), MR000033.
2. Turner, E. H., Matthews, A. M., Linardatos, E., Tell, R. A., & Rosenthal, R. (2008). Selective publication of antidepressant trials and Its influence on apparent efficacy. New England Journal of Medicine, 358, 252-260.
3. Ioannidis, J. P. (2005). Why most published research findings are false. PLoS Medicine, 2(8), e124.

4. Gonon, F., Konsman, J.-P., Cohen, D., & Boraud, T. (2012). Why most biomedical findings echoed by newspapers turn out to be false: The case of Attention Deficit Hyperactivity Disorder. PLoS ONE, 7(9), e44275.
5. Walach, H., & Schmidt, S. (1997). Empirical evidence for a non-classical experimenter effect: An experimental, double-blind investigation of unconventional information transfer. Journal of Scientific Exploration, 11, 59-68.
6. Kennedy, J. E., & Taddonio, J. L. (1976). Experimenter effects in parapsychological research. Journal of Parapsychology, 40, 1-33.