Unsere Meta-Analyse zum Placebo-Effekt ist publiziert [1]

Ich spüre schon seit längerer Zeit einem Phänomen nach, das ich sehr kurios finde: Die Besserungsraten von Patienten in klinischen Studien, die in den Behandlungsgruppen sind und denen, die in den Placebogruppen sind, sind hoch miteinander korreliert, unabhängig von Krankheit und Behandlung. Die Korrelation ist irgendwas zwischen r = .70 und r = .78. Ein typischer Korrelationsplot, den wir vor Jahren in einer ersten Meta-Analyse publiziert haben, sieht so aus:

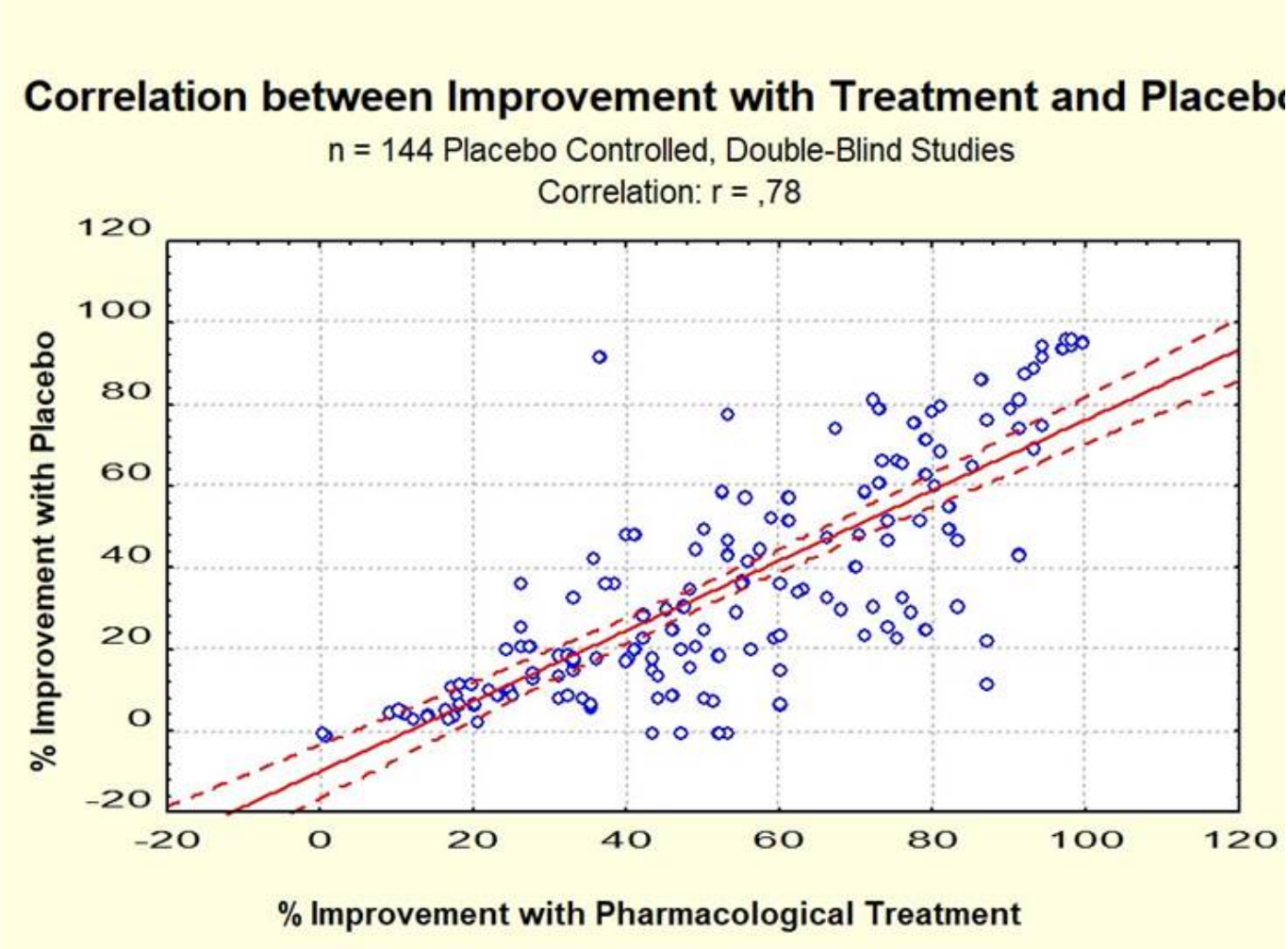

Abbildung 1 – Korrelation der Verbesserung von Patienten in Behandlungsgruppen und Placebogruppen in 144 Studien [2]

Die Vorgängerstudie [2]

Jeder dieser Punkte stellt eine Studie dar. Auf der x-Achse ist die prozentuale Verbesserung

der Behandlungsgruppe abgetragen. Auf der y-Achse ist die prozentuale Verbesserung der Placebogruppe der gleichen Studie abgetragen. Die Datenbasis ist eine Sammlung von 144 doppelblinden, Placebo-kontrollierten Studien irgendwelcher pharmakologischer Substanzen bei irgendwelchen Krankheiten. Gesammelt hat diese Datenbasis seinerzeit Katharina Sadaghiani für ihre Doktorarbeit, die ich betreut habe, nachdem sie schon in ihrer Diplomarbeit einen ähnlichen Befund gesehen hatte. Wichtig dabei ist: Die Sammlung fand systematisch statt, indem sie die vier großen medizinischen Journals damals nach rückwärts nach Studien durchsucht hat, die unseren Einschlusskriterien genügt haben. Diese waren einfach: Langzeitstudien von mehr als 12 Wochen Dauer, pharmakologische Intervention und Placebo als Kontrolle.

Man sieht auf einen Blick: Die Korrelation ist sehr hoch (r = .78). Wann immer eine Studie einen hohen Behandlungseffekt hatte, war der Behandlungseffekt in der Placebogruppe ebenfalls hoch (oder umgekehrt). Wann immer eine Studie keinen hohen Behandlungseffekt hatte, war der Effekt in der Placebogruppe auch niedrig. Dass die Korrelation nicht perfekt war (r # 1.0) ist noch ein beruhigendes Zeichen. Denn wäre sie perfekt, hätte die medizinische Intervention überhaupt keinen positiven Beitrag. Aber dass die Korrelation so hoch ist, ist eigentlich sehr beunruhigend für Fans der pharmakologischen Wirksamkeit. Denn das bedeutet: Der Beitrag des spezifisch pharmakologischen Effektes ist relativ gering. Idealerweise würde man sogar eine Nullkorrelation erwarten, weil sich ja Störeffekte über die Studien hinweg ausmitteln sollten. Tun sie aber nicht. Und interessanterweise kann man diese hohe Korrelation nicht aufklären, indem man unterschiedliche Krankheiten analysiert. Wir haben damals geschlussfolgert: Klinische Studien scheinen ein starkes therapeutisches Ritual darzustellen. Oder aber: Es gibt eine ungeklärte Korrelation zwischen diesen Gruppen, die man u.U. als Verschränkungskorrelation werten muss.

Nun haben wir versucht diesen alten Befund in einer neuen Studie erstens zu replizieren und zweitens noch genauer aufzuklären.

Die Replikationsstudie [1]

In der neuen Studie [1] haben wir also versucht, diesen alten Befund zu replizieren. Wir wollten aber auch die Komponenten des Placebo-Effektes noch näher analysieren. Denn in den Placebogruppen klinischer Studien werden ja gleich mehrere Effekte kontrolliert; das habe ich seinerzeit unter dem Begriff des „Wirksamkeitsparadoxons“ genauer analysiert: [3-5].

Regression zur Mitte

Da ist die sogenannte „Regression zur Mitte“: In jeder Studie, in der zweimal gemessen

wird, am Anfang und am Schluss, taucht eine Tendenz auf, die als „Regression zur Mitte“ bekannt ist. Menschen, die bei der ersten Messung einen sehr hohen Wert haben, haben tendenziell einen niedrigeren bei der zweiten Messung. Das liegt daran, dass die Messinstrumente – Fragebögen, andere Arten der Messungen nicht perfekt sind. Technisch gesprochen: die Reliabilität, die Zuverlässigkeit der Messung ist nicht gleich 1, sondern niedriger. Bei Fragebögen meistens so etwa 0.7 bis 0.8. Das ist ein statistisches Artefakt. Es führt dazu, dass eine Verbesserung suggeriert wird, wo vielleicht gar keine vorliegt. Dieses Artefakt kann man rechnerisch kompensieren, unter der Voraussetzung, dass man die Reliabilität kennt. Das ist bei standardisierten Fragebögen meistens der Fall. Daher haben wir in dieser Nachfolgestudie solche Studien genommen, bei denen immer die gleichen Instrumente verwendet wurden.

Natürliche Tendenz der Krankheit

Es gibt nur wenige Krankheiten, die sich nicht von selber ändern würden, wenn man ihnen unbehandelt ihren Lauf lässt. Die meisten verbessern sich irgendwann mal. Migräne, oder Depressionen, oder Schlafstörungen würden sich häufig auch von selber verbessern, wenn man lange genug beobachtet. Denn die meisten Patienten kommen zum Arzt oder machen in Studien mit, wenn es besonders schlimm ist. Daher ist auch hier die Tendenz gegeben, dass sie sich wohl auch zu einem bestimmten Prozentsatz von alleine bessern würden. Allerdings wissen wir das oft nicht, weil es nur selten untersucht wird. Daher haben wir in dieser Studie auf Krankheiten zurückgegriffen, von denen aus einigen Studien der natürliche Verlauf abschätzbar war.

Fünf Krankheitsdiagnosen mit je 30 Studien

Aus diesen Gründen haben wir 5 unterschiedliche Krankheitsdiagnosen verwendet, bei denen alle mit den gleichen Methoden gemessen wurde und zu jeder Diagnose 30 ähnliche Studien gesucht, wieder systematisch, rückwärts gehend. Bei jeder Diagnose haben wir nur solche Studien verwendet, bei denen eine pharmakologische Intervention untersucht wurde, egal welche.

Die Diagnosen waren:

- Arthrose
- Migräne
- Schlafstörungen
- Depression
- Reizdarmsyndrom (IBS)

Für die Arthrose gibt es ein viel verwendetes Fragebogeninstrument, den WOMAC (Western

Ontario and McMasters University Osteoarthritis Index). Migräne wird mit Tagebüchern erfasst. Für die Schlafstörungen wird häufig das Pittsburgh Sleep Inventory verwendet. Das Reizdarmsyndrom wird mit dem Irritable Bowel Syndrome Quality of Life Questionnaire erfasst und Depression mit der Hamilton Depression Rating Skala. Für unsere Analyse haben wir nur solche Studien verwendet, die diese Instrumente verwendet hatten.

Zu jeder Krankheit hatten wir aus anderen Studien Angaben über die Entwicklung ohne Behandlung, der sogenannten No-Treatment Effekt (NT), und für alle diese Instrumente hatten wir Reliabilitätsmasse. Daher konnten wir

- die Regression zur Mitte und
- die Entwicklung der Krankheit ohne Behandlung

abschätzen und in einer meta-analytischen Regression kontrollieren.

Auch für dieses Studienensemble fanden wir eine hohe Korrelation von r = .73 zwischen Verbesserung unter pharmakologischer Substanz und Verbesserung unter Placebo. Der Korrelationsplot sieht sehr ähnlich aus:

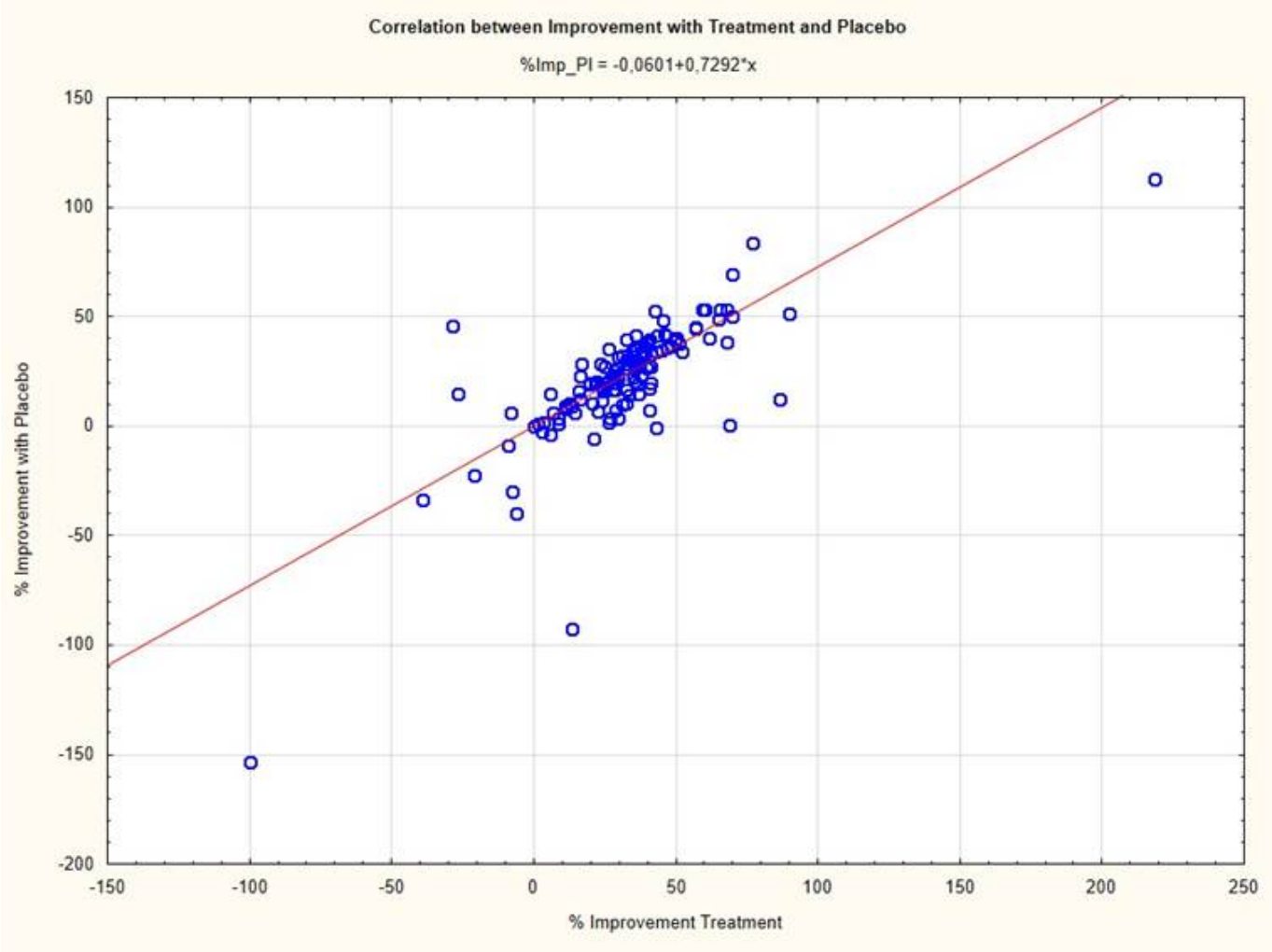

Abbildung 2 – Korrelation zwischen Verbesserung unter Behandlung und Placebo in insgesamt 150 Studien zur Behandlung von Migräne, IBS, Depression, Arthrose oder Schlafstörungen aus [1]; r = .73

Wir sehen also auch in diesem Studienensemble von 150 Studien eine ähnlich hohe Korrelation. Nun ging es uns darum herauszufinden, ob man diese hohe Korrelation aufklären kann und welche Elemente dazu beitragen. Untersucht man die Diagnosekategorien separat, dann sieht man: Die Korrelation ist höher bei Migränestudien und Depressionsstudien (r = .85) und noch höher bei IBS (r = .92) und niedriger bei Arthrose (r = .43) und bei Schlafstörungsstudien (r = .42).

Wir haben mit Regressionsmodellen versucht, Behandlungs- und Placeboeffekte und ihre Varianz aufzuklären. Das gelingt relativ gut und die Modelle können 72 % der Varianz erklären.

Der Behandlungseffekt ist niedriger bei Schlafstudien, bei Multizenterstudien, außerdem spielt die Regression zur Mitte eine Rolle. Aber der Placeboeffekt, bereinigt um die natürliche Tendenz der Krankheit, ist immer noch mit beta = .83 mit dem

Behandlungseffekt korreliert. Anders ausgedrückt: Auch wenn man alle Variablen, die wir miterfasst haben, zur Aufklärung heranzieht, bleibt der Placeboeffekt immer noch die Hauptkomponente bei der Erklärung des Behandlungseffektes.

Umgekehrt kann man auch versuchen, den Placeboeffekt durch eine solche Regression aufzuklären. Auch hier kann man 72 % der Varianz erklären. Der Placeboeffekt ist höher in Multizenterstudien und niedriger, wenn eine Studie mit „Intention to Treat“ ausgewertet wurde. Das ist eine Auswertungsmethode, bei der alle einmal eingeschlossenen Patienten ausgewertet werden, auch wenn sie abgebrochen haben. Aber wiederum spielt der Effekt der Behandlung, bereinigt um den Effekt der natürlichen Tendenz der Krankheit, mit beta = .84 die wichtigste Rolle.

Nochmals anders ausgedrückt: Die Korrelation zwischen Verum und Placebo bleibt bestehen, auch wenn man methodische Artefakte, die natürliche Krankheitstendenz und Studienmerkmale berücksichtigt.

Wir hatten auch die Studienqualität mit einem Instrument, dem sogenannten Detsky-Score, eingeschätzt. Dieser hat gegenüber dem viel verwendeten „Risk-of-Bias“-Instrument der Cochrane Collaboration den Vorteil, dass er einen numerischen Wert ergibt, den man in einer Regressionsanalyse verwenden kann. Und wir sehen: Die Studienqualität hat keinen Einfluss auf die Höhe des Placeboeffektes (oder des Behandlungseffektes) in dieser Studienstichprobe.

Wir können also folgern: Die Korrelation zwischen Verum und Placebo ist kein Artefakt. Placeboeffekte in solchen Studien sind nicht einfach ein Resultat der statistischen Regression zur Mitte, der natürlichen Tendenz der Krankheit sich zu verändern, oder Folge schlechter Studienqualität oder unterschiedlicher Krankheitsentitäten.

Vielmehr ist diese Korrelation robust und Placebo- und Behandlungserfolge in einer Studie sind sehr hoch miteinander korreliert. Vielleicht sind solche Studien einfach sehr potente Heilrituale, so wie alle Heilrituale seit schamanischen Zeiten potent sind und die Selbstheilungseffekte anregen?

Jedenfalls ist die häufig verbreitete Meinung, es seien die spezifischen Wirkungen von Arzneien, die den Hauptanteil einer therapeutischen Wirkung tragen, falsch. Wir können mit unseren Regressionsgleichungen 72 % der Varianz aufklären. Anders ausgedrückt: maximal 28 % des Effektes, vermutlich eher weniger, gehen auf die pharmazeutische Substanz zurück. Der Rest ist die Folge eines Heilrituals.

Ich persönlich habe noch eine andere Erklärung für diesen Effekt, die ich in der früheren Publikation kurz genannt habe [2]: Die Anlage einer klinischen Studie, mit Verblindung und Randomisation, erfüllt die formalen Kriterien, die nötig sind, um eine generalisierte

Verschränkungskorrelation aufzubauen [6-8]. Das würde bedeuten: Ein Teil des therapeutischen Effektes unter pharmakologischer Substanz findet sich in der Kontrollgruppe wieder; das aber nur deshalb, weil durch die Verblindung und die Randomisation eine Verschränkungskorrelation erzeugt wurde. Noch einmal anders ausgedrückt: Man kann eigentlich aufgrund solcher Studien nicht wirklich auf die wahren Effekte von Verum-Interventionen schließen. Vielmehr müsste man sehr unterschiedliche Studientypen heranziehen und durch die Kombination der unterschiedlichen Daten den Effekt herausschälen, wie wir das einmal vorgeschlagen haben [9]. Wenn man diese Effekte wirklich ernst nehmen würde, dann würde der methodologische Kanon ziemlich rasch zerbröseln. Aber sie sind auch der Grund, weswegen bei Interventionen, bei denen die spezifischen Effekte möglicherweise sehr klein sind, eine Trennung von „echten“ und „unechten“ Effekten durch verblindete, placebo-kontrollierte Studien nicht möglich ist.

Quellen und Literatur

1. Schmidt S, Loef M, Ostermann T, Walach H. Treatment Effects in Pharmacological Clinical Randomized Controlled Trials are Mainly Due to Placebo. Journal of Clinical Epidemiology. 2024:111658. doi: https://doi.org/10.1016/j.jclinepi.2024.111658.
2. Walach H, Sadaghiani C, Dehm C, Bierman DJ. The therapeutic effect of clinical trials: understanding placebo response rates in clinical trials – A secondary analysis. BMC Medical Research Methodology. 2005;5:26. doi: https://doi.org/10.1186/1471-2288-5-26.
3. Walach H. Das Wirksamkeitsparadox in der Komplementärmedizin. Forschende Komplementärmedizin und Klassische Naturheilkunde. 2001;8:193-5.
4. Walach H. The efficacy paradox and its consequences for research in psychotherapy (and elsewhere). Psychology of Consciousness: Theory, Research, and Practice. 2016;3(2):154-61.
5. Walach H. The efficacy paradox in randomized controlled trials of CAM and elsewhere: Beware of the placebo trap. Journal of Alternative & Complementary Medicine. 2001;7:213-8.
6. Atmanspacher H, Römer H, Walach H. Weak quantum theory: Complementarity and entanglement in physics and beyond. Foundations of Physics. 2002;32:379-406. doi: https://doi.org/10.1023/A:1014809312397.
7. Walach H, von Stillfried N. Generalised Quantum Theory—Basic idea and general intuition: A background story and overview. Axiomathes. 2011;21:185-209. doi: https://doi.org/10.1007/s10516-010-9145-5.
8. Walach H, von Stillfried N. Generalizing Quantum Theory – Approaches and Applications. Axiomathes 2011;21 (2)(Special Issue):185-371.

9. Walach H, Loef M. Using a matrix-analytical approach to synthesizing evidence solved incompatibility problem in the hierarchy of evidence. Journal of Clinical Epidemiology. 2015;68:1251-60. doi: https://doi.org/10.1016/j.jclinepi.2015.03.027.