



Einsichten eines Wissenschaftsnarren (8)

Wer's glaubt, wird selig!

Heute soll es um den Placebo-Effekt gehen. Wobei wir uns hierbei auch einem weithin unbekanntem Phänomen zuwenden werden: der Regression zum Mittelwert. Und die ist auch für Experimentatoren wichtig.

Die Medizin ist voller Mythen. Manchmal hat man sogar den Eindruck, dass sie hauptsächlich auf Mythen beruht.

Viele dieser Mythen sind so plausibel, dass man ein Narr sein muss, um *nicht* daran zu glauben. Kaum einer zweifelt etwa an der geradezu magischen Effektivität des Placebo-Effektes. Es wird Sie deshalb vielleicht verwundern, dass es recht wenig Evidenz für seine Existenz gibt. Aber einige gewichtige Argumente *gegen* ihn. Die Cochrane-Reviews, immerhin der goldene Standard des systematischen Reviews, konnten keine überzeugenden Belege für seine Effektivität finden. Möglicherweise sind Placebos wirksam bei Therapieergebnissen, die Patienten selbst berichten (*patient reported outcomes*) – insbesondere bei Schmerz und Übelkeit. Allerdings sind die Effekte, sollten sie existieren, wohl recht gering. Keine Wirksamkeit zeigte sich bei sogenannten *observer reported outcomes* – also immer wenn die Studienärzte etwas gemessen hatten.

Weil Sie den Placebo-Effekt für eine der Grundfesten der Medizin halten, und mich für einen Narren, werden Sie jetzt möglicherweise diesen Artikel kopfschüttelnd beiseite legen. Oder Sie geben mir die Chance, Ihnen ein paar Argumente zu liefern, warum es sich hierbei vielleicht tatsächlich um einen Mythos, in jedem Fall aber um ein deutlich überschätztes Phänomen handelt. Sie würden dann auch etwas über die Regression zum Mittelwert erfahren. Dies könnte vielleicht sogar für Ihre eigene Forschung Relevanz haben.

Ein zufällig über oder unter dem Mittelwert ausfallender Messwert wird tendenzi-

ell gefolgt von einem Mess-Resultat, das näher am Durchschnitt liegt. Trivial, nicht? Noch simpler ausgedrückt: Je weiter ein Messwert vom Mittelwert abweicht, desto unwahrscheinlicher ist er.

Der Naturforscher und wissenschaftliche Tausendsassa Francis Galton (1822-1911) hat dies als Erster erkannt – und dem Phänomen auch seinen Namen gegeben: Regression zum Mittelwert. Er nutzte im Jahre 1886 Bevölkerungsregister, um die Körpergröße von Eltern und deren ausgewachsenen Kindern im Erwachsenenalter zu vergleichen. Dabei fand er, dass ausgewachsene Kinder im Schnitt näher an der Durchschnittsgröße liegen, als deren Eltern. Und nur scheinbar paradoxerweise, dass ein großes Kind in der Regel Eltern

»Dummerweise fehlt in fast allen randomisiert kontrollierten Studien eine echte Kontrollgruppe!«

hat, die kleiner sind als es selbst (mehr dazu bei Senn S., *Significance* 8:124-26). Aber was hat das nun mit dem Placebo-Effekt zu tun?

Patient wird man, wenn man Krankheits-symptome hat. Zum Arzt geht man, wenn man diese nicht mehr ertragen möchte oder kann. Der tut dann irgendwas, und zum Glück geht es einem aufgrund der ärztlichen Kunst (scheinbar) nach einer Weile häufig besser.

Oder man geht nicht zum Arzt, sondern weiß selber oder aus der *Apotheken Umschau*, welche Medizin am besten für einen ist (beispielsweise Bachblüten oder Ibuprofen). Nachdem man die Medizin genommen hat, wird es meist nach einigen Tagen besser – und nach einigen Wochen ist der Spuk vorbei.

Voltaire (1694-1778) hat das so formuliert: „Die Kunst der Medizin besteht darin, den Kranken solange abzulenken, bis die Natur die Krankheit geheilt hat“.

Neben der Erklärung der scheinbaren Wirksamkeit von Homöopathie liegt genau hier auch der Hase im Pfeffer beim Placebo-Effekt. Als solchen bezeichnen wir die Verbesserung der Symptome mit einem Scheinmedikament oder einer Scheinprozedur. Die wird, in den *guten* Studien, randomisiert kontrolliert und verblindet mit dem echten Wirkstoff oder Prinzip („Verum“) verglichen. Dummerweise fehlt aber in fast allen randomisiert kontrollierten Studien eine echte Kontrollgruppe! Nämlich eine, die überhaupt keine Behandlung erhält. Nur im Vergleich mit dieser könnte man überhaupt von einem Placebo-Effekt sprechen. Nur im Vergleich mit einer solchen Kontrollgruppe ließe sich klären, wie sich die Krankheit natürlich, also ohne Behandlung entwickelt – und ob Verum- und Placebogruppe überhaupt einen davon abweichenden Verlauf nehmen.

Zum Glück gibt es aber auch solche Studien. Und aus diesen wissen wir, dass der natürliche Verlauf der meisten Erkrankungen fluktuierend ist – und in der überwiegenden Mehrzahl am Höhepunkt der Symptome behandelt wird. An dem Punkt also, wo es ganz natürlicherweise wieder besser wird. Und von hier ab funktioniert Placebo in der Regel *nicht* oder *kaum* – Stichwort „Schmerz, Übelkeit, Stimmung“ – besser als der natürliche Verlauf. Wenig Psychosomatik, viel statistisches Artefakt.

Etwas allgemeiner ausgedrückt kann ein Vergleich innerhalb einer Gruppe zwar zeigen, ob es einem Patienten besser oder schlechter geht – aber nicht, ob und in welchem Ausmaß das auf die Behandlung zurückzuführen ist.

Es kommt aber noch dicker. Die Regression zum Mittelwert versteckt sich in fast allen klinischen Studien und führt dort zur Überschätzung des Behandlungseffektes, egal ob Verum oder Placebo.

Nehmen wir als Beispiel eine Studie, die ein Blutdruck-senkendes Medikament testet. In die Studie wird man aufgenommen, wenn man einen Blutdruck hat, der einen gewissen Wert überschreitet. Rein aufgrund der statistischen Fluktuation werden beim Blutdruck-

messen innerhalb einer Gruppe von Menschen immer welche dabei sein, die bei der Messung einen erhöhten Blutdruck haben, aber keine Hypertoniker sind. Schon bei der nächsten Messung wäre der Wert wieder normal: Regression zum Mittelwert!

Diese Menschen würden aber als Studienteilnehmer aufgenommen, ihr Blutdruckwert vor Behandlung in die Bestimmung des Mittelwerts der Gesamtgruppe eingehen. Nun wird behandelt, die Messung wird wiederholt – und der Mittelwert in der Gesamtgruppe ist jetzt niedriger als vor Gabe des Medikaments. Zwangsläufig wird der Effekt des Medikaments jetzt überschätzt werden, da ja auch die ‚Patienten‘ wieder mitgemessen werden, die gar keinen Hypertonus haben und deren Mittelwert jetzt regrediert ist (ausführliches Beispiel mit Zahlen bei Senn S., *Significance* 8:124-26).

Das wäre alles kein Problem, wenn man jetzt eine echte Kontrollgruppe hätte – also Unbehandelte! Denn auch dort würde man den erniedrigten Blutdruck finden, aber vielleicht nicht so stark wie in der Verum-Gruppe. Leider ist dieser Vergleich aber bei den wenigsten Studien möglich, da eine unbehandelte Gruppe schlichtweg fehlt. In den Studien, in denen eine unbehandelte Gruppe mitge-

führt wurde, fand man keinen Placebo-Effekt – oder allenfalls in geringer Ausprägung bei subjektiven Symptomen wie Schmerz oder Befinden! Was für unser Beispiel heißt: Nicht im Blutdruck.

Vielleicht arbeiten Sie selbst ja mit Zellkulturen oder mit Ratten – und werden deshalb sagen: „Interessant, aber glücklicherweise führe ich ja immer eine Gruppe ohne Behandlung mit. Geht mich also gar nichts an.“

„Vorsicht!“, halte ich jedoch als Narr dagegen. Denn die Regression zum Mittelwert gilt natürlich nicht nur für individuelle Werte, sondern auch für die Ergebnisse von ganzen Studien. Insbesondere wenn sie auf kleinen Fallzahlen beruhen und daher eine hohe Varianz haben, sowie gleichsam über niedrige statistische Power und wenig stringente Signifikanzniveaus von fünf Prozent verfügen. Also die meisten Studien.

Stellen Sie sich also vor, sie machen ein Experiment. Wie immer mit $n=8$. Sie finden einen Effekt, und der ist statistisch signifikant – sagen wir $p<0.03$. Sie sind glücklich. Sie machen noch ein paar andere Experimente für die Studie und schreiben dann das Paper, das den Effekt beschreibt. Wir gratulieren!

Was aber wäre, wenn der eben signifikante Effekt ein falsch-positiver gewesen ist? Und eine Wiederholung des Experiments den Mittelwert in Richtung eines Null-Effektes korrigiert hätte? Also zum Mittelwert regrediert wäre? Durch unsere Fetischisierung von positiven und insbesondere spektakulären (das

»Es ist ziemlich sicher, dass wir häufig falsch-positive Befunde in die Welt hinaus posaunen.«

heißt: *a priori* unwahrscheinlichen) Befunden ist es ziemlich sicher, dass wir häufig falsch-positiven Befunden aufsitzen und diese in die Welt hinaus posaunen.

Das Problem wäre leicht lösbar, aber die Lösung leider wenig populär: Größere Fallzahlen, ausreichende Power, stringente Signifikanzniveaus, Replikationen sowie Publikation auch der negativen und neutralen Resultate. *Good bye, Nature-Paper!* (Siehe hierzu auch den „Wissenschaftsnarr“ in *Laborjournal* 4/2017: 24-25)

Möglicherweise war das alles zu starker Tobak für Sie, und ein bisschen zu viel Mythen-Kritik in weniger als 8.000 Zeichen. Sollten Sie allerdings hinreichend verunsichert sein, um sich noch den ein oder anderen ausführlicheren Artikel zum Thema zu Gemüte führen zu wollen, so finden Sie diese wie immer unter <http://dirnagl.com/lj>.



Foto: BIH/Thomas Rafalzyk

Ulrich Dirnagl

leitet die Experimentelle Neurologie an der Berliner Charité und ist Gründungsdirektor des Center for Transforming Biomedical Research am Berlin Institute of Health. Für seine Kolumne schlüpft er in die Rolle eines „Wissenschaftsnarren“ – um mit Lust und Laune dem Forschungsbetrieb so manche Nase zu drehen.

Draußen kühl?



Drinnen cool!

Ganz frisch seit
1. Januar

” LABORJOURNAL
online

laborjournal.de