

Fallstricke der Meta-Analyse

#### **Description**

# Ein kurzer methodischer Kommentar anlĤsslich einer â??Retractionâ?? unserer Homöopathie-ADHS-Meta-Analyse

Wir hatten uns zu früh gefreut. Ich hatte ja letzten Sommer berichtet, dass wir eine Meta-Analyse zu Homöopathie bei ADHS publizieren konnten, die eine signifikante Effektstärke von g = 0.6 zeigte [1]. Vor Kurzem wurde sie zurückgezogen (â??retractedâ??), und zwar vom Journal, nicht von uns.

Der Hintergrund dazu: Wir hatten einen Extraktionsfehler gemacht, und zwar eine Effektgr $\tilde{A}$ ¶ $\tilde{A}$ ?e positiv kodiert, die eigentlich negativ kodiert geh $\tilde{A}$ ¶rt. Das ist einer der Fallstricke in einer Meta-Analyse,  $\tilde{A}^{1}$ 4ber den ich jetzt selber gestolpert bin. Denn man muss sich immer fragen: Deuten nun die Effekte einer Studie in die Richtung der vermuteten Hypothese, unterst $\tilde{A}^{1}$ 4tzen also die Vermutung, dass der Unterschied f $\tilde{A}^{1}$ 4r die Wirksamkeit einer Behandlung spricht, oder dagegen? In diesem Falle [2] war das Ergebnis nicht nur nicht signifikant f $\tilde{A}^{1}$ 4r die Hom $\tilde{A}$ ¶opathie, sondern wies auch in die andere Richtung. Das h $\tilde{A}$ 2tte in der Analyse mit einem Minus-Zeichen versehen werden m $\tilde{A}^{1}$ 4ssen, was ich schlicht und ergreifend  $\tilde{A}^{1}$ 4bersehen hatte. Und meinen Kollegen ist es auch nicht aufgefallen und so hat sich dieser sehr dumme Fehler eingeschlichen.

Was ist der Effekt eines solchen Fehlers? In der original publizierten und jetzt zur  $\tilde{A}^{1}$ 4ckgezogenen Analyse ist der Effekt  $\tilde{A}^{1}$ 4ber alle sechs Studien g=0.569, mit einer Irrtumswahrscheinlichkeit von p<.001. Das Sch  $\tilde{A}^{1}$ 2tzverfahren ist ein Modell zuf  $\tilde{A}^{1}$ 2lliger Effekte, das der weiten Streuung der Effektst  $\tilde{A}^{1}$ 2rken gerecht wird. Der Effekt war also in dieser Analyse sehr deutlich. Mit dem korrigierten Vorzeichen ergibt sich ein g=0.568 mit dem Modell zuf  $\tilde{A}^{1}$ 2lliger Effekte und ist damit von der Sch  $\tilde{A}^{1}$ 2tzung her sehr  $\tilde{A}^{1}$ 2hnlich. Was sich aber  $\tilde{A}^{1}$ 2ndert, ist die Signifikanzsch  $\tilde{A}^{1}$ 2tzung. Sie ver  $\tilde{A}^{1}$ 2ndert sich zu p=0.053 und verpasst knapp die formale Signifikanzgrenze.

Wenn man nur die vier placebokontrollierten Studien betrachtet, ergibt sich neu eine Effektsch $\tilde{A}$ ¤tzung von g = 0.592, ebenfalls mit dem Modell zuf $\tilde{A}$ ¤lliger Effekte gesch $\tilde{A}$ ¤tzt. Die publizierte Analyse hatte g = 0.605 berichtet, also etwas h $\tilde{A}$ ¶her. Dieser Effekt war signifikant mit p = 0.03. Jetzt ist der Effekt numerisch etwas kleiner; immer noch relativ gro $\tilde{A}$ ?, aber nicht mehr signifikant, n $\tilde{A}$ ¤mlich p = 0.2. Mit einem Modell fester

Effekte w $\tilde{A}$ ¤re dieser Effekt kleiner (g = 0.561), aber signifikant (p < .001). Aber ein solches Modell w $\tilde{A}$ ¤re nicht angemessen, weil die Studien nicht homogen genug sind.

Wir sehen also: Das Vorzeichen hat vor allem einen Effekt auf die Signifikanz der Analyse, weniger auf die Sch $\tilde{A}$ ¤tzung der Gr $\tilde{A}$ ¶ $\tilde{A}$ ?e des Effekts. Das liegt eben daran, dass der Effekt dieser Studie numerisch klein ist, im Vergleich zu den anderen Studien, vor allem im Vergleich zur Langzeitstudie aus Indien, die einen sehr gro $\tilde{A}$ ?en Effekt von g = 1.9 aufweist und die Analyse dominiert. Daher ergibt sich nun durch das negative Vorzeichen dieser einen Studie eine viel gr $\tilde{A}$ ¶ $\tilde{A}$ ?ere Schwankungsbreite, die wiederum die Signifikanzsch $\tilde{A}$ ¤tzung beeinflusst.

Aufgrund dieser gro $\tilde{A}$ ?en Schwankungsbreite ist auch ein Modell fester Effekte unangemessen, auch wenn es signifikante Effekte liefern w $\tilde{A}^{1/4}$ rde.

## Modell fester und zufĤlliger Effekte

Was ist der Unterschied? Bei einer Meta-Analyse legt man immer ein statistisches Modell an die Daten. Ein Modell fester Effekte geht davon aus, dass der wahre zu schĤtzende Effekt der Effekt des Mittelwertes aller Studien plus einer Schwankung, eines Stichprobenfehlers, ist. Diesen schĤtzt man aufgrund der Abweichungen der einzelnen Studien vom Mittelwert im VerhĤltnis zur Anzahl aller Studien, vergleichbar der Definition eines Standardfehlers in der normalen Statistik.

Das Modell zufĤlliger Effekte nimmt nun an, dass zusĤtzlich zum Stichprobenfehler noch eine systematische Schwankung dazukommt, deren wahre GrĶÄ?e man nicht kennt, sondern einfach schĤtzt, mit einem zusĤtzlichen SchĤtzverfahren. Man geht also davon aus, dass die wahren Werte nicht einfach um einen Mittelwert herum zufĤllig schwanken, sondern dass sie zufĤllig schwanken *und* dass noch eine systematische Schwankung hinzukommt. Das ist meistens die realistischere Annahme. Dieses Modell führt in der Regel, vor allem dann, wenn es angemessen ist, zu anderen, oft auch zu gröÃ?eren Effektstärkeschätzungen, aber dafür zu konservativeren Signifikanzschätzungen. Denn die Signifikanz wird hier nicht nur aus dem Stichprobenfehler, sondern aus diesem und der systematischen Schwankung geschätzt.

In den Meta-Analysen, die ich bis jetzt gerechnet und gesehen habe, waren fast immer zuf $\tilde{A}^{\mu}$ llige Effekte passend.

### **Die Retraction**

Das Journal hat vor allem diesen Fehler moniert. Der war in der Tat ein Fehler. Wir hĤtten ihn gerne mit einem Korrigendum verbessert. Das wĤre aus unserer Sicht mĶglich gewesen. Denn es Ĥndert sich an der GesamteinschĤtzung nicht viel. Diese war: HomĶopathie ist vielversprechend, aber die Analyse grļndet auf wenigen und zu weit streuenden Studien und daher sollte man das nĤher untersuchen. Was sich, wie ich gezeigt habe, Ĥndert, ist weniger die EinschĤtzung der GrĶÄ?e des Effekts, sondern die Signifikanz des Gesamtmodells. Und in Sachen Signifikanz gibt es ohnehin sehr unterschiedliche Aussagen. Der Altmeister der psychologischen Methodenlehre, der Harvard-Methodiker Robert Rosenthal, hat mal einen Artikel publiziert, in dem er schrieb â??Surely, God loves the 0.6 as he loves the 0.5â?? [3, p. 1277]. Damit meinte er: Die Fixierung auf ein bestimmtes Irrtumswahrscheinlichkeitsniveau ist reine Konvention und nicht immer klug. Wichtig, das betonte er immer wieder und das hat sich mindestens in der Psychologie durchgesetzt, ist die EffektgrĶÄ?e selber. Dass man diese gegen Zufallsschwankungen absichern muss, versteht sich. Und so kĶnnte man nun sagen: Die EffektgrĶÄ?e Ĥndert sich nicht sehr, aber die EinschĤtzung, wie stark sie eine Zufallsschwankung reprĤsentiert, die Ĥndert sich sehr wohl. Das stimmt. Aber das Ĥndert nichts an unserer EinschĤtzung:

Homöopathie bei ADHS ist auf jeden Fall interessant und sollte weiter untersucht werden. Im Ã?brigen ist mittlerweile eine neue Studie erschienen, die wir in eine verbesserte Analyse aufnehmen werden, die wir dann erneut publizieren werden, diesmal ohne Vorzeichenfehler.

Das Journal hat noch zwei weitere Punkte genannt: dass wir uns bei einer Risk-of-Bias-EinschĤtzung geirrt hĤtten und dass wir bei der EffektstĤrkeschĤtzung der indischen Studie die publizierten EffektstĤrken hĤtten verwenden mļssen und nicht unsere eigene SchĤtzung. Zum letzten Vorwurf kann ich sagen: Das ist aus meiner Sicht falsch, weil die publizierten EffektstĤrkeschĤtzungen der indischen Publikation offensichtlich falsch waren. Warum ist eine andere Frage. Aber ich habe sie anhand der publizierten Daten nachgerechnet und sie sind falsch. Daher habe ich meine errechneten EffektstĤrken verwendet. Zur falschen Risk-of-Bias-EinschĤtzung kann ich nur sagen: Diese hĤngt sehr stark davon ab, welche Information man zugrunde legt. Oft publizieren Autoren nicht alles, was sie gemacht haben, z.B. weil ihnen nicht klar war, dass in 10 Jahren alle Leute nach dieser Information suchen wù¼rden und weil sie Platz sparen mù¼ssen. Wenn man aber weiò, wie die Autoren gearbeitet haben, weil man sie kennt und mit ihnen geredet hat, kann man andere EinschĤtzungen treffen. Man kann darù¼ber streiten, ob das gut oder schlecht, möglich oder falsch ist. Auòerdem sind manche EinschĤtzungen wirklich bis zu einem gewissen Grad sehr subjektiv. Man kann natù¼rlich immer versuchen, auf die sehr konservative Seite zu schwenken. Wenn man das tut, dann ist nichts mehr wirklich gut und verlĤsslich, auÃ?er in sehr wenigen FĤllen.

Der einzige aus meiner Sicht tats  $\tilde{A}^{\mu}$ chlich stichhaltige Fehler, den wir auch sofort zugestanden haben, war also der Kodierfehler. Ob man darauf mit einer Retraction reagieren muss, dieses Urteil  $\tilde{A}^{1/4}$ berlasse ich anderen. Ich pers  $\tilde{A}^{\mu}$ nlich finde, man h $\tilde{A}^{\mu}$ tte auch mit einer Korrektur reagieren k $\tilde{A}^{\mu}$ nnen.

Wenn ich zum Beispiel daran denke, dass die Arbeitsgruppe von Viola Priesemann eine Publikation in Science publiziert hat, die nachweislich und von ihr zugegebener Ma�en mit falschen Daten operiert hat und diese Arbeit nicht zurückgezogen hat [4, 5], dann fragt man sich, mit welchen MaÃ?stäben wer vermessen wird. Wir Homöopathieforscher, weil am Rande stehend, mit sehr scharfem MaÃ?. Eine Arbeitsgruppe am Max-Planck-Institut, die das Lieblingsnarrativ der Regierung bedient, darf schon mal falsche Daten in ihr Modell einspeisen, ohne dass die FAZ nervös wird.

Wer mir das nicht glaubt: Wir haben das alles haarklein publiziert und mit Links nachgewiesen in unserer  $k\tilde{A}^{1}$ /4rzlich publizierten Arbeit in *Futures* [6]. Dort ist auch der Blog von Frau Priesemann verlinkt, wo sie zugegeben hat, dass wir recht haben [5]. Ich kann den Artikel im PDF gerne allen schicken, die sich daf $\tilde{A}^{1}$ /4r interessieren. E-Mail gen $\tilde{A}^{1}$ /4gt.

Was lerne ich draus? Ich werde garantiert keine Daten fÃ1/4r Meta-Analysen nach 20 Uhr mehr kodieren.

### **Quellen und Literatur**

- 1. Gaertner K, Teut M, Walach H. Is homeopathy effective for attention deficit and hyperactivity disorder? A meta-analysis. Pediatric Research. 2022; https://doi.org/10.1038/s41390-022-02127-3.
- 2. Jacobs J, Williams AL, Girard C, Njike VY, Katz D. Homeopathy for attention-deficit/hyperactivity disorder: a pilot randomized-controlled trial. Journal of Alternative and Complementary Medicine. 2005;11(5):799-806. Epub 2005/11/22. doi: <a href="https://doi.org/10.1089/acm.2005.11.799">https://doi.org/10.1089/acm.2005.11.799</a>. PubMed PMID: 16296913.
- 3. Rosnow RL, Rosenthal R. Statistical procedures and the justification of knowledge in psychological science. American Psychologist. 1989;44:1276-84.

- 4. Dehning J, Zierenberg J, Spitzner FP, Wibral M, Neto JP, Wilczek M, et al. Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. Science. 2020;369(6500):eabb9789. doi: https://doi.org/10.1126/science.abb9789.
- 5. Dehning J, Spitzner FP, Linden M, Mohr SB, Neto JP, Zierenberg J, et al. Model-based and model-free characterization of epidemic outbreaks â?? Technical notes on Dehning et al., Science, 2020. . Github. 2020. 6.
- 6. Kuhbandner C, Homburg S, Walach H, Hockertz S. Was Germanyâ??s Lockdown in Spring 2020 Necessary? How bad data quality can turn a simulation into a dissimulation that shapes the future. Futures. 2022;135:102879. doi: https://doi.org/10.1016/j.futures.2021.102879.

**Date Created** 

03.11.2023