Ein paar (weitere) Gedanken zu unserem Methodenartikel über die mangelnde Übertragbarkeit hierarchischen Denkens in der Methodologie

Vor kurzem haben Martin Loef und ich einen Aufsatz im Journal of Clinical Epidemiology publiziert, in dem wir das hierarchische Denken in der Methodologie kritisiert haben [1]. Wir argumentierten, im wesentlichen, dass es kurzsichtig ist, die wissenschaftliche Brauchbarkeit von Befunden nur davon abhängig zu machen, wie intern valide – also wie methodisch stringent – ein Befund ist. Denn, so das Argument, mit steigender interner Validität sinkt in der Regel die externe Validität und damit die Brauchbarkeit des Befundes. Wenn wir nur auf die interne Validität sehen, dann haben wir am Ende viele wissenschaftlich gültige Befunde, die in der Praxis unbrauchbar sind.

Dass interne und externe Validität inkompatible Konzepte sind, haben wir mit noch ein paar anderen Argumenten belegt und gezeigt: dieses Denken führt zu einem Wirksamkeitsparadox, das die mangelnde Stichhaltigkeit der entsprechenden Argumente belegt. Paradoxa weisen immer darauf hin, dass eine Argumentationsweise fehlerbehaftet ist. Wir haben außerdem gezeigt, wie man dem Problem auf den Leib rücken kann: nämlich indem man bei der Zusammenfassung wissenschaftlicher Befunde, bei systematischen Reviews und Überblicksarbeiten, nicht nur die aus randomisierten Studien isolierten Ergebnisse heranzieht, sondern auch die anderen Befunde berücksichtigt. Nur so kann man einen wissenschaftlichen Überblick darüber gewinnen, wie der Erkenntnisstand zu einem bestimmten Gebiet gerade ist. Wir haben das einen „matrixanalytischen Ansatz“ genannt, weil man dabei eine Matrize aufbaut, in der die Zeilen unterschiedliche Studientypen repräsentieren und die Spalten unterschiedliche Befunde, z.B. positive, negative oder unentschiedene. In den Zellen der Matrize könnten auch Effektstärken sein oder eine komplexe Mischung aus Effektstärken und Bayesschen Gewichten. Jedenfalls würden Überblicksarbeiten auf diese Weise stärker das Feld spiegeln als wenn man sich nur auf die randomisierte Studien beschränkt.

Dieser Vorschlag hat eine Diskussion ausgelöst. Mittlerweile haben sich auch Jürgen Windeler, Chef des IQWiG, und Hans J. Trampisch, Chef der Medizinstatistik in Bochum zu Wort gemeldet und ihre Bedenken angemeldet [2,3]. Wenn sich schon so viel beschäftigte Köpfe die Mühe machen zum Diktiergerät zu greifen oder in die Tasten zu langen, dann kann man sich als bescheidener Psychologe geehrt fühlen.

Windeler weist darauf hin, dass unsere Abbildung, in der wir die Hierarchie der Evidenz der EBM charakterisieren und gleichzeitig dazu vermerken, dass mit steigender interner Validität die externe Validität abnehmen würde, insinuieren würde, dass es eine umgekehrte Hierarchie geben würde.

Trampisch fragt, warum man verzerrte Schätzer für Effekte nehmen würde wollen – etwa

aus Studien ohne Randomisation – wenn man Schätzer hat, die verzerrungsfrei sind – eben aus Studien mit Randomisiation. Denn schließlich ginge es bei dem Problem der externen Validität einfach nur um die Transportierbarkeit von Ergebnissen in einen anderen Kontext.

Beide Kommentare gehen von ähnlichen Voraussetzungen aus, die aus meiner Sicht falsch sind. Meine Antworten sind jetzt publiziert [4,5] und bis 6. August über diese Links frei abrufbar:

Hier die Antwort auf Trampisch

Hier die Antwort auf Windeler

Alle Leser können sich mittels der Originale selbst ein Bild machen.

Für die Eiligen kurz zusammengefasst:

Der Disput geht um die Frage: Kann man, in der klinischen Forschung und vielleicht auch anderswo, Inhalt ohne Kontext bestimmen? Anders gefragt: Gibt es eine Messung, die das Gemessene unbeeinflusst lässt? Nochmals anders gefragt: Kann man klinische Effekte, die man in einer Studie findet, einfach nehmen und anderswohin übertragen („Portability“)? Und: Reicht es daher aus, wenn man nur auf die Güte des Inhalts schaut?

Ich meine: nein. Denn der Kontext – eine Studie z.B. – bestimmt den Inhalt mit, nicht ausschließlich und nicht vollständig, aber zu einem Teil. Das bedeutet: was wir in einer Studie sehen, muss nicht unbedingt das sein, was wir in unbeeinflusster Praxis sehen. Was Patienten im Rahmen eines Experiments erleben kann durchaus sehr anders sein als das, was ähnliche Patienten im Rahmen der Routineversorgung berichten. Der Kontext beeinflusst den Inhalt. Das ist, zur Verdeutlichung, auch ein Appell, sich vom klassischen Denken auch in der Methodologie und Pharmakologie zu verabschieden.

Ich habe dafür das Bild von einer Vase geprägt, bei der es eben nicht gleichgültig ist, ob man Rosen oder Lilien hineingibt, sie aufs Fensterbrett oder auf den Tisch oder die Anrichte stellt. Sondern es ist ein bisschen wie mit einer Vase, die nur dann Wasser hält und die darin eingestellten Blumen, wenn es Rosen sind, die man auf den Tisch stellen will, die aber zerbricht, wenn man Lilien auf die Anrichte stellen will.

Genau die Tatsache, dass so etwas im normalen Alltag nicht vorkommt, zeigt, dass Studien und ihre Ergebnisse vielleicht weniger stark normaler Alltag sind, als alle Welt das glaubt. Klar, es wäre wesentlich einfacher, wenn es so wäre. Aber ich glaube, wir dürfen nicht unsere Wissenschaft nach unseren Bedürfnissen einrichten, sondern müssen die Methode der Wirklichkeit anpassen, auch wenn es komplizierter ist.

Quellen und Verweise

[1] Walach, H., & Loef, M. (2015). Using a matrix-analytical approach to synthesizing evidence solved incompatibility problem in the hierarchy of evidence. Journal of Clinical Epidemiology, 68, 1251-1260.
[2] Windeler, J. (2016). There is no reverse hierarchy. Journal of Clinical Epidemiology, 75, 127-128.
[3] Trampisch, H. J. (2016). Portability of therapeutic effects (letter). Journal of Clinical Epidemiology, 75. http://dx.doi.org/10.1016/j.jclinepi.2016.02.030
[4] Walach, H. (2016). Lacking portability of therapeutic effects: the hierarchy of internal validity is insufficient: A response to Hans J. Trampisch. Journal of Clinical Epidemiology, 75, 129-131.
[5] Walach, H. (2016). There is no hierarchy in the first place – A comment on Jürgen Windeler. Journal of Clinical Epidemiology, 75, 128-129.