Some Introductory Thoughts on Understanding Our New Publication

> *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. doi:10.1016/j.jclinepi.2015.03.027*

Before I make a few critical remarks about „Evidence Based Medicine (EBM)“, I feel the need to once again praise this movement for all that it has achieved, what its goals are and what its vision is: medical care based on scientific findings, not on authority opinions; the replacement of eminence by „evidence„. This is an enlightening, liberating impulse, which is very good and important.

Many who have joined this movement, however, do so unthinkingly. They overlook the subtle dialectic that expresses itself in almost all historical processes. That which one thinks one has overcome – here: Authoritarianism – is returning in a somewhat new guise. For what used to be the personal authority of the head doctor, who told us what to do, is now the anonymous authority of the guidelines. And these are mainly fed by already published systematic reviews and meta-analyses. And these, in turn, are almost exclusively based on randomized studies. The rest of the data is simply ignored. That is understandable. Because we humans are lazy creatures and like to take the shortest route and the one where there is the least resistance.

Behind this there is, of course, what at first appears to be a clever idea: study designs can be put into a hierarchical order that reflects the reliability of the conclusions that can be drawn from them. This results in the famous „evidence hierarchy“ of EBM. At the bottom are the simple observations: Case series, well-documented cohorts, but without direct comparison. You have to generate this yourself from your knowledge or from known data. One level higher are study designs with parallel comparison groups, so-called cohort studies, or naturalistic comparisons that have come about naturally. These are, for example, groups of patients who have chosen their treatment themselves and are then compared with each other. Another level up are randomized studies. These are studies in which the experimenter created the groups by allocating patients to the groups using a random code. If you have several such studies, you can combine them in meta-analyses or systematic reviews and have the „true“ knowledge. So far so good.

The theory is now somewhat more sophisticated [1]. Indeed, one can valorize naturalistic studies if they are particularly good or particularly large and devalue randomized studies if they are very small or poorly organized. But the principle and the mindset are the same:

Because this hierarchy is a hierarchy of reliability of conclusions that can be drawn from the studies, technically speaking *a hierarchy of internal validity, that is why you can ignore information from a lower level of the hierarchy once you have better information*. This is why meta-analyses and systematic reviews almost exclusively consider randomized trials. *I have always considered this to be a capital error,* and in this essay we have provided argumentative evidence for this and also shown that and how it can be done differently.

Before we go into a few reasons why I think this mainstream narrative is short-sighted, a reminder is important: one of the founding fathers of Evidence Based Medicine, David Sackett [2], has always emphasized that there are *three* pillars on which EBM rests. 1. the best possible scientific data, 2. the clinical experience of the doctor and 3. the preference of the patient. Very often, only the scientific data is taken into account and the rest ignored.

Now it is important to understand that the *validity of conclusions, the internal validity* of a study, is *only one form* of validity. There is a *different type* of validity that I believe is equally important, *external validity*, often referred to as generalizability. Strictly speaking, there are two more, ecological and model validity, but we ignore them now. External validity indicates whether and for whom the results of a study can be generalized. The mainstream narrative assumes that internal and external validity are linearly additive, as it were, and imply each other: as if internal validity must first be given before one can (or cannot) worry about external validity. We show in the essay that this attitude is factually and logically wrong. I cannot list all the reasons here without rewriting the whole paper. But a few may be mentioned:

1. To generate an experimental clinical trial, i.e. a randomized internally highly valid trial, you have to keep experimental control. One tries to homogenize study groups – by defining and applying inclusion and exclusion criteria. Ideally, this increases the difference between the groups and makes the discriminatory power greater. Technically speaking, one increases systematic variance and reduces error variance. This is, so to speak, implicit in the experimental procedure. The more such inclusion and exclusion criteria are used, the „cleaner" the study becomes. Its internal validity increases. But at the same time, generalizability and external validity decrease.
2. In order to conduct an experimental clinical trial, patients must consent. Many do not. Those who do consent, we know from studies, are not comparable to those who do not consent. So the results don't apply to the patients who don't consent and those who are similar to them.

3. Many randomized trials, especially those for regulatory purposes, limit the treatment time because longer trials are more expensive, and they also limit the patient groups. By both measures, they increase the chance of proving the effectiveness of an intervention and increase internal validity. In doing so, they lose external validity.
4. Randomized trials must, by definition, design their study participants as passive recipients of therapeutic benefits. That may be fine for pharmaceuticals. But it is wrong for all complex interventions that require the activity and cooperation of patients. Therefore, randomized trials can only ever estimate the minimum expected effect size and fail when it comes to producing maximum effects. For this, one needs study designs that give patients freedom of choice and that also leave the option open for practitioners to choose the intervention they are most convinced of.

All of these are domains of „external validity“ or generalizability of study results. This is higher in naturalistic studies, where patients and practitioners can choose what they want to do. In a sense, this hides the other two neglected pillars of EBM in Sackett’s sense, patient choice and physician clinical experience.

Now the point of our argument is: *internal and external validity are not compatible*; they are in some ways mutually exclusive. Any study that increases internal validity decreases external validity, and vice versa. *No study is conceivable, on principle, that increases both external and internal validity together*, nor have I factually seen one where this is the case. Even the reviewers of the paper, to whom I handed over the challenge to show me one, had to pass.

Therefore, internal and external validity are incompatible concepts. And anyone who has studied the formalism of our generalized quantum theory a bit will immediately see that to theoretically model such concepts, one needs a different formalism than the classical, linear-additive one [3]. It follows from this, by the way: the order in which one generates cognition does matter.

Practically, this can be seen in two examples. In experiential medicine there is often a lot of experiential knowledge – applicable, generalizable – which does not have a particularly high methodological validity and was not obtained in an internally valid way. But the knowledge is there. And when hard studies then produce results that are incompatible with this knowledge, this new knowledge is usually ignored.

Therefore, the research logic in complementary medicine is exactly the opposite of conventional research based on pharmacological substances [4]. Conversely, internally highly valid studies often generate good data on the effectiveness of new interventions. But knowledge about their generalizability and applicability often comes much later. Therefore,

dangerous side effects, interactions, or knowledge of very limited usefulness may only be generated after years of approval. In most cases, however, this comes too late. After all, the marketing authorization has been granted, the product is in use and a lot has to happen before the authorities restrict the indication or withdraw the product. As you can see, the *sequence* in which the knowledge was generated plays a role. Technically speaking: *internal and external validity do not* commutate.

What does that mean specifically?

From this, in my view, the following consequences can be deduced:

1. It is not sufficient to use only the supposedly „best“ data, i.e. randomized studies. Their statements must be contrasted with the statements of naturalistic studies.
2. Research methods are not hierarchically related to each other, but rather circular or like in a mosaic: they complement each other [5].
3. Therefore, in systematic reviews, data from all study types should be considered and contrasted in their statements. We have called this the matrix-analytic approach. This involves creating a matrix that maps different study types in the rows and the number of studies or effect sizes that support or refute a hypothesis or are undecided in the columns. Looking at the overall picture, one can quickly see whether there are consistent trends or contradictions. Such contradictions should be resolved, either by taking a closer look at the studies or by considering modifying hypotheses.
4. In any case, the „quick-and-dirty“ method of systematic reviews should stop [6]: I look for the randomized studies and ignore the rest, thinking that I have given a fair, complete and scientifically acceptable overview of the literature. At best, the result is a publication you can show off and then file it away in the bin. Research funders, reviewers, more reviewers, the Cochrane Collaboration – all should stop promoting, writing and publishing such reviews.

Sources

1. Howick, J. (2011). *The Philosophy of Evidence-Based Medicine*. Chichester: Wiley-Blackwell.
2. Sackett, D. L. (1997). *Evidence Based Medicine: How to Practice and Teach EBM*. New York: Churchill Livingstone.
3. Filk, T., & Römer, H. (2011). Generalized Quantum Theory: Overview and latest developments. *Axiomathes, 21*, 211-220.
   Walach, H., & Stillfried, N. v. (2011). Generalised Quantum Theory—Basic idea and general intuition: A background story and overview. *Axiomathes, 21*, 185-209.

4. Fonnebo, V., Grimsgaard, S., Walach, H., Ritenbaugh, C., Norheim, A. J., MacPherson, H., et al. (2007). Researching complementary and alternative treatments – the gatekeepers are not at home. *BMC Medical Research Methodology, 7*(7). www.biomedcentral.com/1471-2288/7/7
5. Walach, H., Falkenberg, T., Fonnebo, V., Lewith, G., & Jonas, W. (2006). Circular instead of hierarchical – Methodological principles for the evaluation of complex interventions. *BMC Medical Research Methodology, 6*(29). http://bmcmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-6-29 Siehe auch: http://www.altmetric.com/details/733156
6. Vickers, A. J. (2010). Reducing systematic reviews to a cut and paste. *Forschende Komplementärmedizin, 17*, 303-305.