Insights DKVF 2024: Preconference and state of the art session

# Problem case evaluation of structural innovations

While Dr Roland Staples (FAU Nuremberg) raised the fundamental question at the preconference seminar of this year's DKVF as to whether the initiators of the innovation fund were using the wrong definition of the term innovation, Prof. Dr Holger Pfaff (IMVF Cologne) and Prof. Dr Jochen Schmitt (ZEGV Dresden) presented their findings. Dr Jochen Schmitt (ZEGV Dresden) presented a proposal for a new evidence-based approach in healthcare policy in the 'State of the Art' session by explaining how to move from 'theoretically best to practically best evidence'.

>> In his presentation, Dr Roland Staples referred to the concept of innovation commonly used in economics, which stands for the (complex) innovations associated with technical, social and economic change that are - and here comes the crux: 'discovered/invented, introduced, used, applied and institutionalised'. In other words, the innovation fund, as it has been implanted in the healthcare system so far, falls short because the ideas recognised as useful or not useful by projects and evaluations alone only become innovations when they are implemented - if they are approved by the Innovation Committee - the main topic of the DKVF 2024 - and are thus available to as many, if not all, insured persons and patients as possible as part of the so-called, but also not precisely defined, standard care.

However, Staples, a research assistant at the Institute of Sociology at FAU Erlangen-Nuremberg's Chair I - Sociology of Work and Industry (Prof. Dr Rainer Trinczek), also addressed a second, equally fundamental misunderstanding - at least from the perspective of organisational and innovation sociology: Innovation is not the same as innovation. According to him, there are both incremental and disruptive innovations. Incremental innovations are usually product innovations (whether these are mobile phones or pharmaceuticals) that can be researched and implemented relatively easily. Disruptive innovations, on the other hand, require the 'figure of a creative disruptor' for their implementation. This person must be able to 'introduce something truly new from the perspective of a company's tradition, but also be aware that such a disruption can also fail. The core statement of the concept of 'creative destruction', which goes back to Schumpeter, is: 'All economic development is based on the process of creative destruction. Through a new combination of production factors that successfully asserts itself, old structures are displaced and ultimately destroyed. Destruction is therefore necessary - and not a system error - so that reorganisation can take place.' (1)

With this excursion into the concept of innovation in economics, which was significantly influenced by Joseph Alois Schumpeter (* 1883 / † 1950), it becomes clear that the Innovation Committee, which is part of the Federal Joint Committee and in which mainly the banks active in the Federal Committee act and determine what is funded and, above all, what is approved and recommended for implementation in standard care and what is not, has a fundamental problem. One that culminates in the following question: Are the stakeholders in the healthcare system, who are dependent on its cash flow, figures of creative destruction?

## Pfaff: 'What is truth?'

Prof Dr Holger Pfaff, Chair of the Brückeninstitut and Director of the Institute for Medical Sociology, Health Services Research and Rehabilitation Science (IMVR) at the University of Cologne, quoted Luhmann, not Schumpeter, to show why innovations take so long to turn a positively evaluated idea into reality. According to Luhmann's systems theory, every society consists of different systems: that of politics, that of the economy, that of civil society and finally that of science. Pfaff: 'They are all self-referential, so-called autopoietic systems with their own code.' Politics is about power, the healthcare industry is about money, civil society is about satisfying needs and science is about nothing less than the truth. According to systems theory, however, the four systems do not listen to each other. Pfaff: 'They hardly influence each other, that is Luhmann's prediction. That's a bit sobering.' According to Pfaff's own thesis, all systems also want to preserve their structures: the professionals want to carry on as before, the politicians want to leave everything as it is and the self-administration anyway. According to Pfaff, this means: 'The systems are fundamentally resisting the influence of science.' And if we also introduce doubts ourselves, this leads to a double structural conservatism. This leads to either dispensing with evaluations altogether or politicians saying 'if I don't hear anything from you, I'll decide over a glass of wine in the evening'. 'At least that's what one health minister said,' added Pfaff.

And with regard to the congress theme of the DKVF 2024, he explained: 'I'll be quite critical: implementation science develops techniques so that civil society and the healthcare industry finally adopt the things for which we have discovered the truth. And then we are disappointed when that doesn't happen because the systems say: 'I'm not interested in that.' And even more critically: 'They actually want to force the evidence-based form into society. You can also ask yourself: where is the participation?'

To make matters worse, even the truth is a question of point of view, which means that there can be several. For example, when it comes to the question of whether colonoscopy is able to reduce mortality, there are two truths: the relative and the absolute reduction. Pfaff: 'This alone shows that there can be several truths, even where we have data.'

According to Pfaff, there are also several levels of truth. This becomes particularly apparent when basic scientists - whether statisticians or epidemiologists - who are concerned with nothing less than the total truth, assess health services research studies. Taking the much-discussed WiZen study (2) as an example, according to Pfaff, all those who say that on the basis of this study it is not

yet possible to say whether the intervention is really good are right from their respective basic scientific point of view. This apparent contradiction is mainly due to the fact that health services research is an 'applied science' whose task it is to 'ensure that reasonably well-evaluated interventions are put into practice'. Pfaff continues: 'Because this is the case, we health services researchers do not have a view of the totally pure truth.'

Although health services research is of course always about the highest achievable truth, according to Pfaff, this conceals the difficult issue of evidence levels, or more precisely, the difference between the theoretically best evidence and the practically best evidence. And the problems and misunderstandings caused by this discrepancy.

According to Pfaff, the declared aim of health services research - to promote rational health policy on the basis of evidence-based knowledge - has so far been achieved by applying the principles and methods of evidence-based medicine. This in turn is based on randomised controlled trials (RCTs) and meta-analyses of as many RCTs as possible; preferably the 1a or 1b level of evidence and therefore the theoretically best evidence. Anything below this is rejected by IQWiG, although there are now movements to soften this procedure.

Or publicly criticised, as was the case with the WiZen study (3). In his editorial in the Deutsches Ärzteblatt entitled 'Evaluability of the effect of certifying oncology centres', Andreas Stang argued that the study under discussion (4) was below the level that could normally be accepted.

Similar to Staples, Pfaff explained that innovation is not the same as innovation and that a distinction must be made between three types of innovation, even when an evaluation is being planned:

- Product innovations (medicines, drugs)
- Process innovations (guidelines etc.)
- Structural innovations

However, the EBM system, which according to Pfaff is of course justified, is not suitable for every type of innovation. An evaluation via RTC is of course the case for product innovations, but only to a very limited extent for structural innovations. Using the example of innovation fund projects, he explained that a pilot should actually be carried out three to three years before the start of the project in order to test the basic principle to be introduced. Ideally, a four-year study should then be carried out if the MASI framework were to be taken seriously, and only then should the project application follow, which would then take a total of eight years. The results then have to be published and reviewed, which could take another seven or eight years. Pfaff: 'That means we are 15 years too late.' But only if no replication of the study results is required, which is virtually impossible in the case of structural innovations because all established structures have different contexts and therefore different outcomes. Pfaff's first interim conclusion: 'We have a real problem: we need a solution for structural innovations.' There is only one choice: 'We either leave it at that or wait 20 years.' Whereby the latter is not an option, because new technologies have been developed in the meantime or others are already outdated and the structures themselves are also changing. In other words: 'There are several problems that stand in the way of the evaluability of structural innovations.'

Firstly, there is what was already described in 1948 in the essay 'Resistance to Change'. The reason for this is that people do not always do what we as scientists imagine, which at the very least contaminates the experiment. Secondly, the large amount of time required ideally leads to the pilot being dispensed with, thereby integrating 'teething troubles' into the project study applied for and co-testing it, which in turn can falsify the result. As a complete package of interventions is always tested in such complex package tests, it is not known afterwards which components are really effective. This in turn also makes replication more difficult.

All of this leads to high innovation and evaluation costs for structures, which would be a deterrent. The situation is different with drugs, where the 'trial and error' principle can simply be applied. Pfaff: 'This is precisely the great advantage of RCTs: I don't need a theory. You can test an idea by going down the trial and error route and seeing until you eventually find the solution. You can't do that with structures, also because it takes far too long.' RCT results are also still valid after 20 years because they assume a stable environment, a drug that does not change and a person who also does not change.

In view of dynamic environments and rapidly outdated structural innovations, the EBM system does not work. All of this, but above all the limited applicability of the criteria of evidence-based medicine, brings structural innovations to the 'limits of evaluability', according to Pfaff. Pfaff: 'For this reason alone, no matter how hard we try, our evaluation results can always be criticised.' There will always be a critic who wants the pure truth. They would not do this out of malice, but 'because they have learnt that the truth' is paramount. Even if - another trip to Luhmann - there is no such thing as the absolute top truth.

The fatal thing is: 'This sows doubt.' He quoted from the book 'Merchants of Doubt' (5) by Erik M. Conway and Naomi Oreskes: 'It only takes one or two people to express doubt and the majority will no longer believe in scientific results. According to Pfaff, this in turn has an impact on implementation: 'If some people start not doing something because of even very minor doubts - which may also be genetically implanted - healthcare research will either only give cautious advice or no advice at all, because we can't be sure whether an efficacy is really fully proven.'

In his view, this in turn leads to a scientifically induced structural conservatism. His definition, which can be read in the article 'Give innovation a chance: Overcoming structural conservatism of evidence-based medicine and health policy decision-making by returning to Sackett', soon to be published in the journal 'Health Care Research & Implementation': 'A practice that makes decisions in a way that systematically - intentionally or unintentionally - stands in the way of the development and implementation of new structures and thus indirectly preserves old structures.'

Pfaff's (self-)admonition: 'We are partly to blame for the fact that nothing moves forward because we express too much doubt about our own studies. This is our part in the structural conservatism that is responsible for the fact that the development and implementation of new structures is only progressing slowly and old structures are being indirectly preserved.' And this is against the backdrop that the old structures have never been tested because they were created in their own time and their possibilities. Pfaff: 'But we are imposing an evaluation quality on new things that the old structures didn't even have to pass because they simply came into being that way.'

**Doppelter Strukturkonservatismus:**

Der wissenschaftliche Strukturkonservatismus trifft auf Systeme (politische und gesundheitswirtschaftliche Systeme), die sich nicht verändern wollen und selbst strukturkonservativ sind

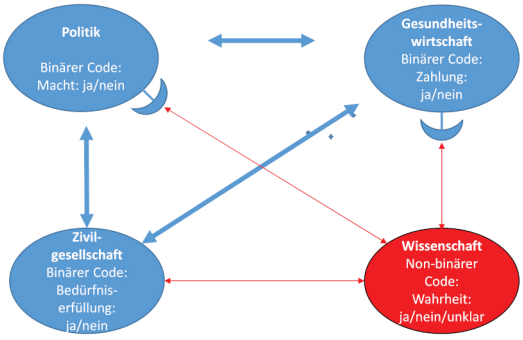



***Fig. 1:*** *Double structural conservatism. Source: DKVF 2024 presentation 'State of the Art: From Theoretical Best to Practical Best Evidence: A Proposal for a New Evidence-Based Approach to Health Policy' (Holger Pfaff, Jochen Schmitt, Ibrahim Demirer).*

In other words, says Pfaff: 'We are under pressure to act.' Health services research can't want the gut to decide in the end because 'we don't deliver'. This also means, Pfaff continues: 'Before we no longer provide any evidence at all, we have to find a middle ground between the highest and none at all.' In future, we will therefore have to differentiate between three types of evidence:

(a) the theoretical best evidence, which EBM strives for with its implicit freedom from doubt memo.
b) the best available evidence, which David Sackett describes in his definition
c) the practically best evidence that achieves the highest level of evidence under the given circumstances, which Pfaff and his colleague Schmitt bring into play.

This means nothing other than that the maximum possible level of evidence must first be defined for each condition if the latest methods are used. Conversely, the definition of maximum evidence also requires the will of the research community to adhere to this maximum: 'If it's possible, we'll do the study; if not, we won't accept it.'

This is not about pragmatic evaluation or even realistic evaluation. It is about nothing less than striving for the 'highest achievable evaluation design under the given contextual conditions on the basis of the latest evaluation research under the influence of mechanistic studies'. For this work, Pfaff and Schmitt propose a structured consensus-finding process modelled on the S3 guideline development process: 'It must be determined what constitutes the highest achievable practical evidence in the given context.'

So far, it is only clear that the best available level of evidence will certainly differ depending on the context and the given political, economic, social, ethical, psychological, organisational, regional, cultural, technical and structural (framework) conditions. It may be that an RCT can indeed be conducted in context A, but not in other contexts, but perhaps a controlled, quasi-experimental, quantitative or correlational study can.

Nevertheless, according to Pfaff, it remains undisputed that there will always be an uncertainty gap between the theoretical best evidence and the practical best evidence. However, this uncertainty gap must be accepted by science, politics and society, knowing full well that it can be exploited by critics at any time. According to Pfaff, politicians must also accept that they have to make decisions under uncertainty. Pfaff: 'EBM is - to exaggerate - something for scaredy-cats.' In the healthcare system, as in the economy, you have to learn to make decisions under uncertainty. However, the uncertainties must be extremely minimised by the scientific community trying the maximum possible and pushing the boundaries methodically.

Referring to the WiZen study, which in his view represents the best available evidence for the tested constellation, he added: 'It may be possible to do even better, but we don't yet know how.'

## Schmitt: the question of 'how'

Prof. Dr Jochen Schmitt, Director of the Centre for Evidence-based Healthcare (ZEGV) at the Carl Gustav Carus University Hospital and Medical Faculty in Dresden, and currently Deputy Chairman of the DNVF, explained the 'how' in his lecture 'EbMplus - international perspectives and developments'. Schmitt's comments should be preceded by the observation that the definition of the 'EbMplus' model, which was developed by the two German healthcare researchers Pfaff and Schmitt, emerged in parallel with the international 'EbM+' development. Both 'EbM+' and 'EbMplus ' - this is very important according to Schmitt - come from the actual EbM movement. It is therefore not a countermovement, but rather an expansion of the traditional EBM concept with the help of the integration of additional dimensions and a broader and more diverse evidence base.

According to Schmitt, EbM+ (international) is significantly characterised by two strands of development:

**Prof Jon Williamson,** Professor of Reasoning, Inference and Scientific Method, Leeds, with his central theme of causality: 'Evidence of Mechanisms should be integrated with evidence of correlation to better assess causal claims'. He postulates that a causal conclusion requires a 'correlation claim' and a 'mechanistic claim', whereby one type of evidence is not sufficient to demonstrate causality. Williamson also introduces 'evidencial pluralism' in reference to Bradford Hill - a model that, according to Schmitt, ultimately amounts to a combination of quantitative, qualitative and mechanistic studies. According to Schmitt, he is saying that evidence based on one study and one design is not sufficient to demonstrate causality. It is therefore always necessary to combine qualitative studies with quantitative and mechanistic studies in order to overcome the efficacy-effectiveness gap. According to Williamson, effectiveness is 'efficacy plus external validity'. Schmitt: 'This is highly interesting because this external validity arises via mechanisms that are as robust as possible and always work in the same way under different conditions.'

**Prof Trisha Greenhalgh,** Professor of Primary Health Care, Oxford, with her approach of 'real evidence-based medicine'. According to Schmitt, she calls for a reform and re-evaluation of traditional EBM in order to adapt it

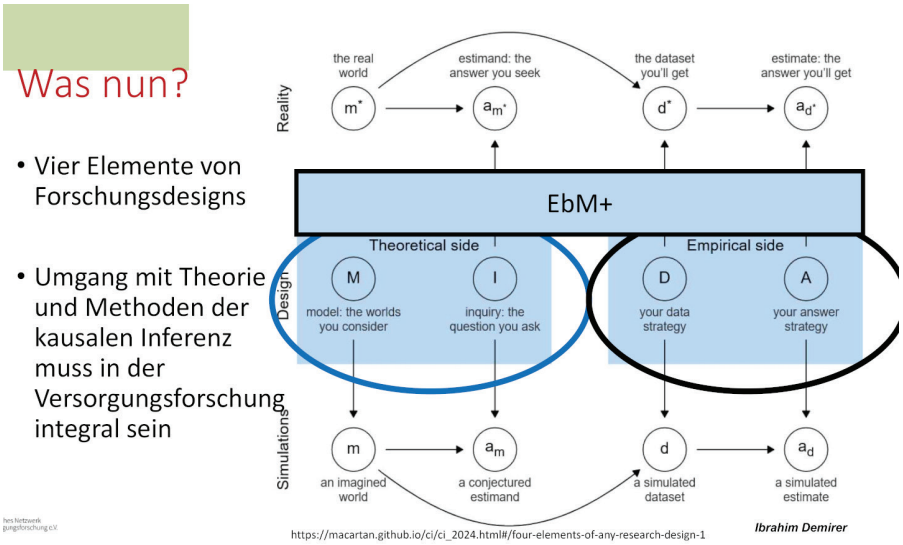



***Fig. 2:*** *Four elements of research designs. Source: DKVF-2024 lecture 'EbMplus - international perspectives and developments' (Jochen Schmitt).*

to the challenges of a complex and dynamic healthcare landscape, which requires interdisciplinary approaches: medicine, social sciences, informatics, exchange of knowledge and experience.

It is important to note that the international understanding of 'EBM+' means that only empirically proven mechanisms and only empirical, data-backed theories count. And secondly, that '+' primarily stands for biomedical correlations. This is different with the German 'EbMplus' , where the 'plus' primarily refers to the application and inclusion of social science theories.

This difference is important because - according to Schmitt - 'causality in complex interventions can be context-dependent' and therefore probabilistic mechanisms with their technical, interpersonal and psychological components always play a role alongside deterministic ones, which will always have an influence on the outcome.

Small note: Pfaff and Schmitt also initially differentiated between deterministic and probabilistic theories and then focussed more on probabilistic theories because they believe that these are decisive, especially for everything beyond the 'simple interventions' scenario.

Another important difference is that international 'EbM+' development has so far primarily focussed on 'micro-level interventions', as Schmitt put it. The German 'EbMplus' model, on the other hand, focusses primarily on process and structural innovations. According to Schmitt, this means:

'Although German health services research has only made a few references to international EbM developments to date, we can also make a relevant contribution internationally!'

In addition to pure application research, German health services research must once again focus more on basic research. According to Schmitt, fundamental and central principles and mechanisms need to be developed and tested in order to work on understanding mechanisms in a healthcare and interpersonal context. <<

by: MVF Editor-in-Chief Peter Stegmaier

**Literature**

1: https://t1p.de/ois0v
2: https://t1p.de/idijm
3: https://t1p.de/74mwm
4: https://t1p.de/df0o9
5: https://t1p.de/bipnt

**Links to the presentations given at the DKVF::**

Pfaff: https://t1p.de/zfhms
Schmitt: https://t1p.de/61sgt

Pfaff's and Schmitt's comments are based on their consecutive papers:
doi.org/10.3389/fpubh.2021.727427
doi.org/10.1111/jep.13890
DOI: 10.1055/a-2350-6435

## Citation

Stegmaier, P.: 'Problem case evaluation of structural innovations', in „Monitor Versorgungsforschung“ (06/24), S. 16–19. http://doi.org/10.24945/MVF.05.24.1866-0533.2663