

Shifting from Theoretical Best Evidence to Practical Best Evidence: an Approach to Overcome Structural Conservatism of Evidence-Based Medicine and Health Policy

Von der theoretisch besten Evidenz zur praktisch besten Evidenz: ein Ansatz zur Überwindung des Strukturkonservatismus in der evidenzbasierten Medizin und Gesundheitspolitik



Authors

Holger Pfaff¹, Jochen Schmitt²

Affiliations

- 1 University of Cologne, Faculty of Human Sciences & Faculty of Medicine and University Hospital Cologne, Institute of Medical Sociology, Health Services Research and Rehabilitation Science, Chair of Quality Development and Evaluation in Rehabilitation; Cologne, Germany
- 2 Zentrum für Evidenzbasierte Gesundheitsversorgung, Med. Fakultät der TU Dresden, Universitätsklinikum Carl Gustav Carus Dresden, Dresden, Germany

Key words

evidence-based medicine, health policy, structural conservatism, innovation, uncertainty, decision making

Schlüsselwörter

Evidenzbasierte Medizin, Gesundheitspolitik, Strukturkonservatismus, Innovation, Unsicherheit, Entscheidung

published online 2024

Bibliography

Gesundheitswesen

DOI 10.1055/a-2350-6435

ISSN 0941-3790

© 2024. The Author(s).

This is an open access article published by Thieme under the terms of the Creative Commons Attribution-NonDerivative-NonCommercial-License, permitting copying and reproduction so long as the original work is given appropriate credit. Contents may not be used for commercial purposes, or adapted, remixed, transformed or built upon. (<https://creativecommons.org/licenses/by-nc-nd/4.0/>).

Georg Thieme Verlag KG, Rüdigerstraße 14, 70469 Stuttgart, Germany

Correspondence

Prof. Jochen Schmitt
Universitätsklinikum Carl Gustav Carus an der Technischen Universität Dresden
Zentrum für Evidenzbasierte Gesundheitsversorgung
Fetscherstraße 74
01307 Dresden
Germany
jochen.schmitt@uniklinikum-dresden.de

ABSTRACT

There is disparity in the healthcare sector between the extent of innovation in medical products (e. g., drugs) and healthcare structures. The reason is not a lack of ideas, concepts, or (quasi-) experimental studies on structural innovations. Instead, we argue that the slow implementation of structural innovations has created this disparity partly because evidence-based medicine (EBM) instruments are well suited to evaluate product innovations but less suited to evaluate structural innovations. This article argues that the unintentional interplay between EBM, which has changed significantly over time to become primarily theoretical, on the one hand, and caution and inertia in health policy, on the other, has resulted in structural conservatism. Structural conservatism is present when healthcare structures persistently and essentially resist innovation. We interpret this phenomenon as an unintended consequence of deliberate EBM action. Therefore, we propose a new assessment framework to respond to structural innovations in healthcare, centered on the differentiation between the theoretical best (possible) evidence, the practical best (possible) evidence, and the best available evidence.

ZUSAMMENFASSUNG

Im Gesundheitswesen ist ein Missverhältnis zwischen dem Ausmaß an Innovationen in Bezug auf medizinische Produkte (z. B. Arzneimittel) und dem Ausmaß an Innovationen in Bezug auf Versorgungsstrukturen festzustellen. Dies liegt nicht daran, dass es an Ideen, Konzepten und (quasi-)experimentellen Studien zu

Versorgungsorganisationen mangelt. In diesem Beitrag stellen wir vielmehr die These auf, dass wir es mit einer schleppenden Implementierung von Strukturinnovationen zu tun haben und dass dies zum einen daran liegt, dass das EbM-Instrumentarium zwar sehr gut zur Evaluation von Produktinnovationen geeignet ist, aber nur bedingt zur Bewertung von Strukturinnovationen. Wir zeigen in diesem Aufsatz, dass sich aus dem ungewollten Zusammenspiel aus einem über die Zeit deutlich veränderten, nun vor allem theoretischen EbM-Verständnis und gesundheitsspolitischer Vorsicht und Trägheit ein systematischer Strukturkon-

servatismus resultiert. Strukturkonservatismus ist gegeben, wenn die Strukturen des Gesundheits- und Versorgungssystems ein starkes Beharrungsvermögen aufweisen und gegenüber Innovationen weitgehend resistent sind. Dieses Phänomen interpretieren wir als eine unintendierte Folge absichtsvollen EbM-Handelns. Um auch bezogen auf Strukturinnovationen in der Gesundheitsversorgung handlungsfähig zu sein, schlagen wir einen neuen Bewertungsrahmen vor, in dessen Mitte eine Differenzierung zwischen der theoretisch bestmöglichen Evidenz, der praktisch bestmöglichen Evidenz und der bestverfügbaren Evidenz steht.

Introduction and problem definition

A guiding principle of modern healthcare policy is the design of the healthcare system as a learning system [1, 2] oriented toward the triple, quadruple, and quintuple aims of healthcare [3–6]. The concept of a learning healthcare system presupposes that product, process, and structural innovations are successfully evaluated using EBM methods and then rolled out and practiced broadly until further evaluations indicate that emerging forms of care have surpassed existing forms. However, Germany and other countries have shown that while a learning system works well for product innovation, it is less effective for process innovations and almost ineffective for structural innovations [7–9].

This implementation deficit has several explanations, such as slow acceptance despite the high benefits, external disruptive factors or sluggish behavioral changes due to habits [9–12]. However, this article proposes an additional explanation: We argue that the slow implementation of structural innovations results from EBM methods and instruments being well suited to evaluate product innovations (e. g., new drugs) but less suited to evaluate structural innovations. Hence, we demonstrate that systematic structural conservatism results from the unintended interplay between a significantly evolved, primarily theoretical understanding of EBM and health policy caution and inertia. Structural conservatism exists when healthcare structures persistently and essentially resist structural innovations. Structural conservatism is an emergent phenomenon, not merely the result of a movement seeking to preserve the existing healthcare order and the prevailing interests and power relations.

This structural conservatism results in structural innovations continuing to struggle against established structures, even if they are highly likely to be effective. This article highlights the role of science in this problem and proposes various solutions. We advocate for a more robust relationship between science, politics, and practitioners in the healthcare system, making it possible to formulate scientific recommendations and political decisions more rationally and pragmatically [13]. In times of multiple crises and challenges in the healthcare system (e. g., skilled worker shortages, demographic changes, and digital transformation lags) that urgently require reorganization and further development at the structural and organizational levels, taking action in healthcare policy is imperative based on the highest practicable level of evidence.

The scientific principle of the best evidence: Its meaning, origin, and areas of application

The original EBM called for the “best available evidence,” not the highest evidence – it was strictly application-oriented because its founders were practitioners [14]. A critical goal of these first-generation EBM advocates was to view clinical epidemiology and EBM as resources for applying evidence in treating patients [15]. This “*application-oriented EBM*” aims “to achieve the integration of research results in clinical practice;” hence, “EBM proposes a formal set of rules to help clinicians interpret and apply evidence” [16]. Clinicians make numerous shared decisions with patients daily while directly responsible for these choices. In clinical patient care, decisions are always active processes. For instance, the decision to maintain or not to initiate a therapy must be actively, directly justified, and communicated to the patient.

In contrast, today’s second EBM-generation representatives of “*theoretical EBM*” are no longer practitioners (e. g., NICE and IQWiG) and are not confronted with the need of making immediate, pragmatic decisions. They do not have to compromise and can always request the theoretically best evidence adhering to pure EBM methodology. They also bear no direct responsibility for decisions based on their rules. The first EBM generation’s originally pragmatic and enabling EBM concept was expanded by the second generation into a theoretically pure concept with the highest scientific standards. These standards and criteria are appropriate from a very theoretical, basic research perspective, which exclusively seeks for an absolute truth about nature and its functioning irrespectively of any practical consequence or application in practice. This second generation advocates a concept that could be described as “pure EBM,” or “theoretical EBM,” requiring that studies provide the highest theoretically possible level of evidence [17, 18]. One example is the requirement that strong evidence-based recommendations be based on meta-analyses of several double-blind, comparative randomized parallel group studies with narrow 95 % confidence intervals [19–20]. In contrast to the original EBM concept, with decisions based on practical clinical experience, patient preference, and study evidence, this second generation of pure theoretical EBM advocates base all recommendations on the principle of theoretical best evidence from studies, irrespectively of the context and related possibility to achieve this theoretically best possible level of evidence from studies.

Sociologically, the development of EBM can be interpreted as the process of an idea that became independent, during which a separate functional subsystem developed with its own institutions (e. g., NICE and IQWiG). This functional subsystem emancipated itself from the “parent” system (i. e., clinical practice), largely sealed off from external influences, as consistently observed with such independent functional subsystems, so-called autopoietic systems [21].

Ergo, the originally pragmatic idea has become so fully independent that “theoretical EBM” is applied indiscriminately to all possible innovations (product, process and structural innovations). This one-size-fits-all approach fails to consider that the principle of theoretical best evidence no longer applies in certain constellations and framework conditions, as is often the case with structural innovation. Thus, practicable (clinical decision-making), multidimensionally integrated (with internal evidence, external evidence, and patient preference), and dialectical principles have been transformed into principles far removed from practice that almost dogmatically, always reflexively, and without contextual reference demand the theoretically best level of evidence. This shift, while understandable in the context of basic research, ultimately leads to an inability to act when the theoretically best level of evidence cannot be provided.

Indeed, these crucial differences between first-generation and second generation EBM advocates have not yet been critically addressed in the scientific community as possible reasons for many decision-making dilemmas in today’s healthcare system.

Whether in its original or current form, EBM is a well-established method for reducing complexity (e. g., a randomized controlled trial (RCT) reduces multi-causal relationships to one factor) and uncertainty in recommendations and decision-making. Ideally, applying the EBM approach effectively reduces the uncertainty in decision making. Hence, EBM serves politics as a reliable provider of truth and an eliminator of doubt since randomization eliminates confounders. Nonetheless, the external validity and generalizability to real-world settings are low.

The first EBM generation addressed the problem of low external validity by integrating the practitioner’s experience and contextual knowledge with patient preference. However, neither concept factors into today’s “pure EBM.”

At its core, EBM that demands the theoretical best level of evidence strives to reduce residual uncertainty for scientists and decision-makers to a theoretical minimum by utilizing appropriate study design and methods to rule out all alternative explanations (i. e., high internal validity). Thus, second-generation EBM proponents primarily support the basic sciences undergirding EBM, particularly clinical epidemiology, which uses statistical and biometric methods to describe causal relationships between exposures and (health-related) conditions/results/outcomes.

The observation that epidemiology and statistics are essentially basic instead of application-oriented sciences [22] is central to our argument. The basic sciences generally strive for pure knowledge and causal truth and must apply the entire arsenal of methods and procedures to arrive at pure knowledge. In the case of interventions, these methods and procedures converge in the criteria of the theoretically best possible evidence.

Notably, representatives of these basic disciplines must criticize a study that does not satisfy the highest form of evidence on purely academic, theoretical grounds. Thus, they legitimately move within their scientific system, striving for absolute truth focused on critical appraisal instead of practical decision-making. As a result, the guiding role model of the second-generation EBM community is the critical methodologist, not the decision-maker in politics and medical practice. In this case, the autopoietic system of basic science is unconcerned with the consequences on the political system and practice.

One consequence (usually unintentional) is the inhibition of innovation in the healthcare system. Hence, albeit unintentionally, basic science supports (or even propagates) the healthcare system’s structural conservatism, resulting in the “unanticipated consequences of purposive social action” [23] and the “unintended consequences” in complex systems [24–26]. Unlike basic scientists, applied researchers (e. g., health services researchers) must think more broadly to consider the consequences of a strategy of scientific purity [27] by moving from disciplinarity to transdisciplinarity.

The principle of theoretical best evidence fails when it comes to structural innovations

Product innovations can be tangible (e. g., new medications and medical aids) or intangible (e. g., health apps) and usually consist of a product core, the product exterior as perceived by the customer and various additional services [28]. Process innovations are innovations in healthcare procedures. They exist at the macro level (e. g., the care pathway for strokes), the meso level (e. g., internal hospital treatment pathways for strokes), and the micro level (e. g., the organization of participatory decision-making for breast cancer).

Structural innovations in the healthcare system can also be found at the macro level (e. g., introducing levels of care and replacing specialist wards with care groups), at the meso level (e. g., mandatory quality standards and the concentration of care in certified centers), and at the micro level (e. g., changes in decision-making structures). Structural innovations are defined as novel changes in the organizational and operational structure of a healthcare provider that have not yet been implemented by the provider [29]. One example at the macro and meso levels is the nationwide introduction of certified cancer centers [30–35].

Distinguishing between these three innovation types is relevant for our considerations because they correlate with the ability to fulfill the criteria of the theoretical best level of evidence. For example, product innovations are often ideally suited to applying the principles of the theoretically best evidence. Ideally, the evaluation of product innovations, especially pharmaceutical innovations, can be conducted so that all criteria required for the theoretically best level of evidence are fulfilled. Thus, statisticians, biometricians, and epidemiologists no longer have objections since all doubts about the innovation’s effectiveness are eliminated. However, when evaluating structural innovations, the principle of the theoretically best evidence reaches its limits in various respects, as will be shown below.

Limited manipulability

A core feature of RCTs (Randomized Controlled Trial) and CRTs (Cluster-Randomized Trial) is the manipulability of the independent variable [36, 37] using an arbitrarily planned intervention [37]. In reality, however, structural innovations as an independent variable can only be manipulated to a limited extent. Structural innovations change care structures at the macro level, organizational structures at the meso level, and interaction structures at the micro level. The principle of manipulability is particularly applicable at the micro level and poses less of a problem than at the meso and macro levels. Thus, experimental manipulability is subject to strong, pragmatic limits at the macro and meso level.

Resistance to change

Resistance to change involves forms of collective resistance to planned structural changes. More than processes and products, structures are linked to interests, resources, and power [9] and can trigger conflicts of interest and power struggles resistant to change [38, 39]. Indeed, structural innovation is not isolated but includes existing structures to be supported, supplemented, or replaced by them. Since stakeholders (e. g., employees and shareholders) have the most at stake with replacement innovations, resistance to change is widespread among them [9]. Resistance to change also occurs when there are no power interests at stake, e. g. when those affected want to hold on to their habits, routines and safety precautions for supposedly good reasons.

Assembly costs and time

Moreover, even if all stakeholders are willing to change, structural innovations cannot be flexibly manipulated at will (e. g., flipping a switch). Building structures takes time and money [9] and can sometimes require several years since existing care structures must either be transformed or new care structures established. The financial costs of converting or establishing supply structures add to the time and energy required. Hence, these singular events have high material and immaterial costs and require significant time (e. g., compared to animal or psychological experiments).

Reversal costs: Imaginary and real

Beyond initiating and activating structural innovation, dismantling an unsuccessfully evaluated structural innovation cannot be neglected after an experiment is conducted. While a drug can easily be discontinued, a structural innovation cannot, especially when creating new organizational units, buildings, facilities, apparatuses, or investments such as developing personnel and organizations. Thus, anticipating a possible reversal of structures after the (negative) evaluation must be considered from the outset. Nonetheless, if a structural experiment has a negative outcome, the self-interests of those who want to hold on to the given situation (e. g., employees wanting to keep their jobs) can often prevent the necessary reversal.

An even more critical case occurs when a structural innovation is not to create a new structure but to abolish an old one (e. g. closing a rural district hospital and replacing it with an outpatient healthcare network). Indeed, the closed district hospital cannot be reopened if the experiment ends negatively. Hence, structural in-

novations cannot generally be established and dismantled as product innovations can be.

Complexity

A further limitation to evaluating structural innovations at the highest level of evidence, as required by second-generation EBM advocates [19], refers to complex interventions. EBM, initially developed for individual treatment decisions in a clinical context, is suitable at the population level using simple, stable interventions (e. g., drugs, medical devices, and patient training) to minimize decision uncertainty by applying randomization at the individual level. However, complex interventions involving several influencing factors, actors, system components, and interactions limit the applicability of theoretically best EBM standards. This limited evaluability is because complex medical innovations often affect several systems (e. g., technical, physical, psychological, and social systems). When dealing with systemic contexts, positive or negative side effects within and between systems (whether intended or unintended) must be considered to obtain a holistic picture of the consequences of a particular decision and to assess the overall impact of an isolated individual decision.

Environmental dynamics and EBM lag

The principle of the theoretical best evidence generally reaches its limits in dynamically developing fields of application, as shown below with two typical cases.

The first case deals with completely new threats (e. g., COVID-19), where EBM is unsuitable for providing knowledge to decision-makers at the outset [40]. As a result, third-generation EBM representatives have recently attempted to accelerate knowledge generation and systematization through rapid reviews and living guidelines [41–44], which was labelled the “organic turn” of EBM [45] or “pragmatist turn” [46]. However, accelerating processes cannot address the fundamental dilemma of the EBM lag [47]. The EBM lag is the period between an innovation’s emergence and the availability of systematic reviews of RCTs and meta-analyses on its effectiveness related to a specific primary outcome [47].

The second case is when the technologies used in a new care structure develop rapidly [48–50]. For example, if a care structure is developed today using ChatGPT 4.0, a systematic review of this AI-engineered structure may only be available a decade later. By then, however, ChatGPT may be available in version 10.0 or discontinued and replaced by a qualitatively different form of AI, thus rendering the systematic review obsolete. Since digital technologies become outdated quickly, the knowledge gathered on them also becomes outdated. Therefore, the digital transformation exposes a new weakness of EBM concerning product innovations such as digital health applications (DIGA) and other health technologies.

Both cases above are affected by the EBM lag, which is implicitly a topic in some articles on EBM [50, 51]. In the middle of the last century, Ogburn noted that culture (e. g., legal regulations) frequently lagged behind technological progress (i. e., the cultural lag). We define EBM-lag as the time that elapses between the emergence of the care innovation and the publication of systematic reviews, meta-analyses and (living) guidelines on the effectiveness of this care innovation. As mentioned, acceleration and flexibility attempts cannot fundamentally change this lag [45].

Limited randomizability of structural innovations and low evaluation culture

EBM is a valuable approach to generating evidence in an evaluation culture where structural innovation researchers and practitioners are willing to engage in cluster-randomized experiments [52] and stepped-wedge designs [53]) to gain long-term benefits through increased knowledge. However, the core problem in this context is the randomization of individuals, medical practices, clinics, regions, and (federal) states to an intervention or control group. Randomization is *the* core element of the principle of theoretically best evidence, which controls for known and unknown confounders that no other method can match [50]. However, randomization is challenging if not impossible to apply to structural innovations, especially in non-government healthcare systems.

The first boundary of randomization occurs often in a low evaluation culture [54]. In this case, the subjects refuse to be the object of the study (i. e., acting as “guinea pigs” in an experiment), the second boundary results from the aforementioned high costs and expenditures of material and time involved in creating a structural innovation. In this case, an innovative care structure is introduced in a care organisation because it is part of an experiment and not because an independently conducted strategy development process within the care organisation has led to the conclusion that this innovative care structure is the right structure for the future (“not invented here”-problem). In an experiment, the free decision on designing the future care structure is replaced by a decision from outside (e. g. science). This creates resistance to experimentation.

Hence, the manipulability of the independent variables is limited since few social units (e. g., districts, countries, organizations, and clinics) will take on this effort if they are “only” randomly part of the control group (which, according to the research hypothesis, typically results in worse outcomes than the intervention group). Randomization is also difficult or even impossible if already established structural innovations exist. In healthcare, a historically evolved structure exists in which randomization is no longer possible and selection effects have already occurred. Although good evaluation designs can be used in a historically evolved care structure, they never reach the theoretically best evidence level to satisfy the pure EBM criteria of the theoretically best evidence.

The constitution of the healthcare system

As explained above, the theoretically best level of evidence reaches its limits in the case of expensive, complex structural innovations that affect power and interests. These innovations cannot be evaluated regularly according to the theoretically best evidence principle in non-governmental healthcare systems. Structural innovations can only be properly evaluated if their further application is halted until a prototype is tested and permitted only after a cluster-randomized study is conducted to confirm or disprove their effectiveness. Here, structural innovations require compulsory randomization.

One example where a randomized design is impossible is the introduction of care groups instead of the current specialist wards, as currently discussed in the context of the German healthcare reform [55]. In Germany, the ideal would be a three-arm randomized study that randomly assigns individual federal states to either the

“64 care groups”-concept (structural intervention A), the “128 care groups”-concept (structural intervention B), or the current specialist ward concept (the control group) [55]. According to the Cochrane risk of bias tool [56], the study would also have a high risk of bias since it could not be conducted in a blinded/masked manner vis-à-vis the federal states and the population.

According to EBM maxims, such an “ideal” approach would only be conceivable in centralized, state-run systems, not in decentralized, market-based, liberal health systems. Voluntary randomization would also be possible in decentralized healthcare systems for meso-level interventions, for example, at the organizational level (hospitals). However, it would entail several other bias issues (e. g., due to the motivation and willingness to change) in addition to the blinding/masking problem.

Structural conservatism as an unintended consequence of applying the principle of theoretically best evidence

Ergo, the evaluation of structural innovations generally cannot meet the requirements of the second EBM generation for the theoretically best level of evidence. This perspective and attitude are related to Popper’s [57] falsification principle and justified by the classical test theory [58]. Using conventional significance tests [59] researchers primarily aim to avoid type 1 errors (false positives)[60]. This scientific approach is inherently conservative since the old should only be replaced by the new if the new fulfills all criteria of indubitability with the alternative hypothesis accepted.

As a consequence new structures are not introduced, nor are old ones abolished due to statistical concerns, even if it is highly probable (but not highly certain) that the new is better than the old. This structurally conservative approach of “pure EBM” in politics and practice leads to science systematically disadvantaging the new. Therefore, type 2 errors become more likely. Some arguments have suggested that this type of error (a false negative) can be far more problematic than type 1 [60]. As Fiedler and colleagues (2012) state, “*We show that the failure to assertively generate and test alternative hypotheses can lead to dramatic theoretical mistakes, which cannot be corrected by any kind of rigor applied to statistical tests of the focal hypotheses*” [60]. The most unintended consequence of this precautionary strategy for healthcare is that structural conservatism is supported by science, if not (co-) generated by it in the long term.

Indeed, the classical theory of significance testing is necessary for drugs if one does not want to risk that a new drug is worse than a verified old drug. However, this principle is only applied to drugs if there are good, well-tested alternatives, as is usual in modern medicine for all common diseases, but not for Orphan drugs targeting patients with rare diseases without effective approved treatments.

In the case of structural innovations, the falsification principle combined with the approach of “pure” EBM means that structural innovations are inevitably classified as uncertain or doubtful in effectiveness. Consequently, they do not receive strong recommendations. This outcome is welcomed by interest groups in the healthcare sector, who cling to the status quo for various reasons, such as power interests, preserving resources, inertia, and change fatigue [9]. It allows them to embrace their existing structures with-

out changing things. This result would not necessarily be negative if the existing system also underwent an empirical test before its introduction into healthcare. However, healthcare structures usually evolve historically to “prove” themselves without being evaluated as rigorously as their introduction into practice according to “pure” EBM. This problem is specific to structural innovations since product innovations (e. g., new drugs) can usually be compared with an existing alternative tested for effectiveness. However, with structural innovations, the previously untested is compared with the new, the latter of which must meet the highest statistical requirements. Hence, structural interventions are at an inherent disadvantage compared to product innovations.

This mechanism is not without consequences. For example, suppose a political necessity and urgency exists to adapt care structures to new social circumstances. In politics, no attempt is usually made to systematically derive scientific recommendations and political and practical decisions from study evidence. Rather, decision-makers tend to trust experts or make political decisions directly and independently – without resorting to experts or study evidence – so that politics remains capable of acting while demonstrating its ability to act [44, 61]. Examples include adjustments to care structures and processes during the COVID-19 pandemic [44] and discharge management regulations. Health policy, pressured to make decisions, often acts independently of evidence-based science, usually due to a lack of or controversial top-level evidence. In this case, policymakers act without evidence-based knowledge of “pure EBM.” In our view, however, there must be a compromise between option A (a successful but non-evidence-based decision) and option B (failed attempt to make a decision due to doubts about whether one has reached the theoretical best evidence). We will propose such a compromise below.

Measures to overcome structural conservatism

From the viewpoint of applied health services research, we present a catalog of measures that build on each other and can interrelated as a program. Together, they form the basic framework of a program of evidence generation and interpretation to guide action for structural innovations.

We propose the following algorithm for this program:

1. raising awareness of methodological trade-offs and deriving consequences for a study design/program and the strength of recommendation,
2. defining a priori the theoretical best and the practical best level of evidence with regard to the structural innovation and related context,
3. conducting a (rapid and/or scoping) review and integrating state-of-the-art theory to determine the best available evidence,
4. presenting the difference between the best available and the practically best achievable level of evidence, including the respective uncertainty of the decision, and
5. confronting decision-makers with the (evidence) situation and jointly agreeing on a research program.

6. A) If an evidence-based decision is desired and sufficient time is available, then the research program should be completed and the results should be presented to the political decision-makers.
- B) If there is no time to carry out the research program, then a transparent decision should be made considering the best available evidence, the state of theory, and a modeling/ impact analysis.

Step 1: Raise awareness for methodological trade-offs and derive consequences for study design and strength of recommendation

Various conditions at the macro and meso levels of a healthcare system can be more or less pronounced depending on the system. Therefore, they require compromises regarding the highest achievable level of evidence. The third chapter in this article describes some of the most important framework conditions, including resistance to change and the lack of acceptance of randomization. Legitimate reasons for necessary compromises and the resulting methodological compromises need to be standardized in the research community a priori and communicated. ► **Table 1** overviews this process. The list in ► **Table 1** is not exhaustive but is merely intended to stimulate discussion.

Step 2: Defining a priori the theoretical best and practical best level of evidence

We propose distinguishing between three levels of evidence:

- theoretical best evidence (theoretical best possible evidence),
- practical best evidence (practical best possible evidence), and
- best available evidence (currently highest available evidence).

The theoretical best possible evidence refers to the highest evidence level achievable in an ideal experimental world regardless of the circumstances, with no levels of evidence above it. The highest level of evidence requires no alternative explanation for the empirical result apart from the “intervention” factor.

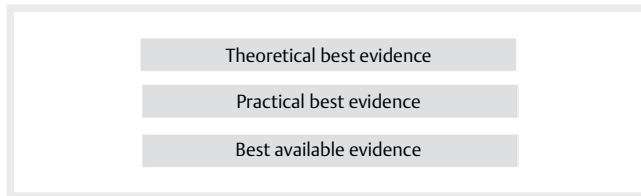
The theoretically highest level of evidence is achieved when a metaanalysis of several double-blind, comparative (cluster) randomized parallel group studies with a narrow 95 % confidence interval is available for a structural innovation [19, 20]. It must be defined in each case and visualized as an ideal target to determine the methodological steps taken in an ideal world where the power of the factual does not exist to recommend a decision with the lowest uncertainty.

In contrast, the practical best evidence is the highest achievable evidence under specific circumstances and conditions, including political, economic, social, ethical, psychological, legal, data protection, and organizational conditions, which define the limits of the study design. These limitations can result in the practical best possible evidence being far away from the theoretically best evidence, which is regularly the case with structural innovations. For example, when dealing with a structural innovation in a state healthcare system, the practical best evidence can be moved more toward the theoretical best evidence than in a liberal healthcare system because it can essentially be ordered. Nevertheless, even state healthcare systems experience clear limits to the evaluability

► **Table 1** Barriers and challenges to generating the theoretically highest evidence and consequences for study design and decision uncertainty.

Barriers/challenges to generating the theoretically highest evidence (selected cases)	Example	Resulting limitation/ risk of bias (examples)	Proposed strategies to best cope with barriers/challenges in study design (examples)	Consequences for interpretation, uncertainty assessment, and decision-making
Complexity of intervention	Peer review in diagnosis and therapy (e.g., multidisciplinary tumor boards)	The causal outcome effect of different interventions cannot be determined.	Assessing the quality and completeness of the intervention components while describing possible relationships to the overall outcome qualitatively	The uncertainty effect of complex intervention has not increased, and the causality of single components cannot be determined (but may not be relevant)
Instability of intervention (over time)	Digital health innovations (based on AI-methods, involving updates as technology advances)	The evaluated intervention is not valid anymore at the time of the systematic review.	Organic turn strategies & agile EBM: rapid reviews, using high-quality theories, and EBM + studies	Accepting preliminary uncertainty; using a scenario method to project possible future scenarios
The latency between the introduction and the full effect of the intervention	Quality assurance programs in hospitals	Misclassification bias (e.g., due to partly introduced intervention) → Bias toward the null effect	Allow a run-in phase to ensure the full effect of the intervention under investigation (may be several years)	The true effect of the intervention may be larger than the effect assessed in the study; monitoring the effects over time of study and later; a mix of summative and formative evaluations
Randomization is impossible for ethical reasons.	High likelihood that intervention is superior (e.g., initial cancer treatment in certified centers ¹)	Confounding due to known and unknown determinants of outcome related to the intervention (but not in the causal pathway)	Considering all possible/known confounders on patients, providers, and regional levels in a study design and analysis	Considering the likelihood of residual confounding or confounding by unmeasured determinants to qualitatively change the study results and related recommendations for practical decision-making
Randomization is impossible for cultural and political reasons.	Meso- and macro-level interventions in decentralized healthcare systems			
Randomization is impossible for legal reasons.	Nurse-led hospital wards in Germany			
Impossible masking of the intervention	Generally applies to meso- and macro-level structural interventions	a) Contamination of the control group → Bias toward the null effect (H0) b) Hawthorne effect? ² (vanguard effect) → Bias toward H1	Reducing competitive efforts in the control group, using a historical control group, using CRT, using strategies to reduce the Hawthorne effect	The true effect of the intervention may be a) larger than the effect assessed in the study or b) smaller in the long run (from the vanguard to routine effect)
Data protection regulations make valid assessment of relevant outcome data impossible.	Cause-specific mortality related to healthcare structure interventions in Germany	The effect of the intervention on this outcome cannot be assessed.	Considering available surrogate and additional outcomes	Basing recommendation on surrogate outcomes; (e.g., in the long run, changing data protection regulations)
Limited power due to limited observation units (applies to macro-level interventions)	Reorganization of healthcare planning on the federal and state levels	False negative study result	Applying simulations/modeling methods based on study results and high-quality theories	If the effect size is moderate to large while simulations and high-quality theories support effectiveness, then strong recommendations may be given even if statistical significance cannot be determined.

¹[31, 63, 64] ²[65–66]



► **Fig. 1** Distinguishing three critical levels of evidence.

of structural innovations when striving for the highest theoretically achievable evidence.

The theoretically best evidence and the practically best evidence should never refer exclusively to one study. The studies conducted (e. g., innovation experiments) must be subjected to replication and also pass replication tests (s. ► **Fig. 1**).

The best available evidence is used in this context per Sackett [62], focusing on policymakers rather than physicians. Therefore, evidence-based health care planning is the conscientious, explicit, and judicious use of current best evidence in making practical and political health management and policy decisions about the care of specified patient groups, thereby integrating collective clinical and healthcare expertise with the best available external clinical and healthcare evidence from systematic research.

Depending on the state of research, this level of best available evidence may be closer or further from the practically best (possible) evidence and the theoretically best (possible) evidence. The best available evidence is a) frequently available to scientists and decision-makers in a systematic review or evidence synthesis and b) can be made available in an emergency to scientists and decision-makers after scanning published evidence. Therefore, despite its potential weakness, this best available evidence is better than an ad hoc political decision without evidence.

Step 3: Conducting a (rapid and/or scoping) review and integrating state-of-the-art theory

When urgently needed political adjustments arise in response to rapidly changing conditions, the best available evidence should be processed using rapid and/or scoping reviews to scan and record available evidence. These reviews should also help determine whether reviews are available on the status of a) mechanistic studies concerning the question and their content (the status of EBM + ; [67–69]) and b) theoretical work in this field [70].

The application of the EBM + theory approach with the three phases a) using theories to identify theoretically causal mechanisms and to plan interventions (phase 1), b) using EBM + to empirically identify causal mechanisms and further specify the intervention (phase 2) and c) conducting high-level EBM studies with regard to the intervention (phase 3) can be helpful here [45]. In this process, theories first provide orientation about the healthcare world to reduce complexity and highlight the possible side effects of structural innovations, thereby allowing starting points to derive accurate interventions while helping explain phenomena and their relationships.

The second step is the EBM + approach, “which systematically considers mechanistic evidence (studies which aim to explain which

factors and interactions are responsible for a phenomenon) on a par with probabilistic clinical and epidemiological studies” [69]. When uncovering mechanisms of effect, non-randomized procedures and methods should also be used if they contribute to uncovering existing causal mechanisms and pathways [69].

In the third step, mechanistic studies are supplemented by (cluster) randomized experimental studies to apply the classic EBM approach [45]. Reviews on reviews of these three steps can help quickly prepare the best available evidence for decision-makers.

Step 4: Presenting the difference between the best available and the practical best achievable level of evidence

Many health services researchers will want to claim that their study has produced the best possible evidence (i. e., the most practically achievable evidence). Others will want to disagree because they apply different standards. One solution to this potentially emerging problem is to define the best possible evidence in the scientific community by consensus in advance, separately for each typical constellation of framework conditions, because each constellation of innovation type and framework condition has a valid, practically highest level of evidence (practical best possible evidence). Therefore, the criteria for the practically highest level of evidence should be defined in advance for each innovation type and typical condition to ensure a consistent subsequent assessment of the available evidence while avoiding conflicts of interest in interpreting the study results. An exemplary general guiding question is: What is the best possible level of evidence for structural innovations in a non-government healthcare system with restrictive data protection regulations? The answer to this question depends on the context, the structural intervention and the specific research question. ► **Table 1** provides some strategies to cope with given limitations to provide the theoretically highest evidence for different scenarios.

Importantly, this best possible level of evidence should be determined by a legitimized group that should clarify the typical constellations of conditions and how to define the best possible evidence for each. This best possible evidence should be defined in advance as a guiding objective (e. g., by research funders) so that projects can be oriented accordingly in their study planning.

When a priori defining the catalog of requirements for the practical best evidence for a typical constellation of framework conditions (i. e., practical conditions), comparing it to the theoretical best evidence (under ideal conditions) would be desirable. The differences between the two evidence categories would help specify the underlying conditions of the necessary compromises (e. g., data protection, data availability, no acceptance/possibility of randomization, and too few observation units). These specifications could (politically) justify the need to change these conditions.

Step 5: Confronting decision-makers with the (evidence) situation and jointly agreeing on a research program

In the case of the applied sciences, which are not primarily concerned with finding the absolute truth, striving to achieve the highest possible level of evidence is only warranted if political decision-makers are willing to make an evidence-based decision.

Therefore, we propose another solution for healthcare research based on applied science: publicly acknowledging residual uncertainty in the findings while still making clear recommendations. The basis for this reasoning is the commitment of science and healthcare research to recommendations under uncertainty. Thus, this proposal requires abandoning the pursuit of ultimate certainty while being bolder with recommendations. The objective could be to determine the likely benefits and harms of a structural intervention and compare them with the benefits and harms of the status quo. Thus, based on a previously defined scientific process, the (unavoidable) residual risk associated with maintaining existing structures would be communicated to decision-makers, including the residual risk associated with introducing new structures. In return, political decision-makers should transparently explain why they decided for or against the new structures. Hence, a politically accountable, active decision should always be made, even if the status quo is maintained, since a non-decision is also a decision based on a variation of Watzlawick's [71] theorem.

Managing residual risks

Residual risks accompany a positive or negative recommendation for introducing a structural innovation, despite the difference between the theoretical best evidence and the best available evidence that ultimately forms the basis of the recommendation. The risk scientists face is that they may “wrongly” recommend. However, the more evidence is available for the recommendation, the lower the risk of error. In contrast, the higher the residual risk, the greater the uncertainty of the political and practical decision-makers, so the risk of making the wrong decision grows. Again, taking a small residual risk is better than no decision in the pursuit of maximum certainty; it is also superior to taking a very high risk because a politician under pressure to act often decides without an evidence base.

Ergo, health services research, as a basic research-oriented applied science, must conduct internal residual risk management regarding practical recommendations. For policy advice, this procedure entails providing data and evidence with instruments, methods, and theories that enable decision-makers to know and assess the residual risk of a positive or negative decision.

Systematic monitoring

An essential element of risk management in making recommendations and decisions is recording the possible consequences of political and practical decisions through the “systematic monitoring” of the effectiveness and impact of (structural) healthcare innovations in the sense of continuous evaluation. Models exist for such monitoring processes in the public health sector [72]. Thus, healthcare-related data can be used sensibly and beneficially [73, 74].

As with residual risk management, systematic monitoring accepts uncertainty while committing to formulating the best possible evidence as a pragmatic goal to strengthen the recommendation. Monitoring is appropriate and sensible for any deviation from the highest theoretically achievable level of evidence, yet it is indispensable when the best available evidence deviates from the practically best achievable evidence. Monitoring must focus on aspects emerging from comparing the best available and best possible evidence.

Systematic monitoring can allow science, politics, and practice to learn a posteriori the extent and consequences of the residual risk. Furthermore, these stakeholders can also learn how to better assess future residual risks. Above all, monitoring helps to fine-tune after learning processes are completed so that the negative consequences of a (minor) wrong decision can be quickly identified, rectified, or mitigated. The combination of EBM and learning-based medicine is the best basis for the emergence of a “learning healthcare system”.

Structural innovation assessment

Impact monitoring can be systematized by integrating it into a systemic impact assessment. The basic principles of systems thinking should be considered via systemic analysis of the effects of the introduction and non-introduction of any measure by determining the intended and unintended complex consequences of a deliberate action [70, 75–79]. These consequences could be presented as “if-then” causal relationships to explore the various chains of effects and their cross-relationships in detail in three steps:

- a) analysis of the main effect by implementing an intervention-related causal analysis about the selected main effect (primary outcome),
- b) analysis of side effects by conducting an intervention-related causal analysis about the intended and unintended side effects and immediate and long-term consequences, and
- c) translating the communication of “if-then-knowledge” about the main and side effects to decision-makers in politics and practice utilizing a structural innovation assessment report similar to the former technology assessment reports ([80]. These reports should be far more systemic and broadly aligned than many health technology reports [81, 82]. However, this undesirable narrow focus often characterizes more recent HTA reports [83].

Step 6: Policy and practice decide situationally and flexibly under the guiding principle of the highest achievable evidence

We distinguish between two possible situations: there is a) enough time or b) not enough time to prepare the decision based on the best possible evidence.

Scenario A: When an evidence-based decision is sought with sufficient available time, the best possible research program should be completed with the results presented to political decision-makers. They would then explain how the available best possible evidence was incorporated into their decision-making. If it were not incorporated, these political decision-makers would be obliged to state the specific reasons and substantiate why it was not used and a different political decision was made. In this case, the scientific community – particularly the EBM community – would have to accept this decision as political. Nevertheless, the scientific community would inform the decision to the best of their knowledge and in good conscience (i. e., an informed political decision). In this case, scientists would have done their “jobs” well and reached their power’s justified limits. However, health policymakers would then be responsible for the main and side effects identified in the structural innovation assessment report if they occurred as predicted.

Scenario B: With insufficient time to determine the best possible evidence, a decision should consider the best available empirical evidence and theories with a modeling analysis to predict the most likely impact of (i) implementing and (ii) not implementing the structural innovation. The residual risk in the decision and the associated decision uncertainty should be transparently identified.

CONCLUSION: RETURNING TO SACKETT WHEN DEALING WITH STRUCTURAL INNOVATIONS

Our considerations have shown that with evidence-based health policy advice for structural innovations, we should distance ourselves from the ultimate demand for the highest level of theoretical evidence, i. e. the theoretical best evidence. Instead, we should strive for Sackett's maxim by combining the best available evidence with experience and knowledge about the object (in Sackett's case, the patients, while in our case, the healthcare system with its framework conditions) to make the best policy decision. This "back to Sackett" principle requires that active decisions MUST be made in health policy, similar to how doctors (ideally with patients) make decisions daily. Even deciding not to change care structures must be an active political decision, and the consequences must be ascribed to health policy decision-makers.

However, this principle must be linked to aspiring to advance the best available knowledge – if it is not yet at the level of practical best evidence – toward the best possible knowledge defined in advance (i. e., toward the highest practically available level of evidence). Both should be aligned with the guiding principle of the theoretically best achievable level of evidence. Indeed, a de facto unattainable level of theoretically best achievable evidence may be the guiding principle but not a specific benchmark. Instead, the objective is to allow the innovation a reasonable chance of realization. Since a decision favoring innovation requires accepting uncertainty, systematically monitoring its effectiveness concerning main and side effects is imperative (i. e., did side effects occur, and were they intentional or unintentional?). Depending on the results, practice and health policy can take countermeasures. Based on such an approach, implementing a learning healthcare system that relies equally on evaluation and monitoring is possible, thus maintaining a deliberate course between structural conservatism and innovative ventures.

Conflict of Interest

Prof. Pfaff has received institutional funding for scientifically-initiated research from the G-BA/Innovation Fund and the Federal Ministry of Education and Research (BMBF). He chaired the expert advisory board of the Innovation Fund (G-BA/Federal Ministry of Health) and is on the scientific advisory board of the BGF Institute of the general local health insurance fund (AOK) and co-editor of the health report of the company health insurance funds (BKK).

Prof. Schmitt reports institutional grants for investigator-initiated research from the German Federal Joint Committee, German Ministry of Health, German Ministry of Research, European Union, German

Federal State of Saxony, Novartis, Sanofi, ALK, and Pfizer. He participated in advisory board meetings as a paid consultant for Sanofi, Lilly, and ALK. Prof. Schmitt serves the German Ministry of Health as a member of the German National Council for Health and Care.

References

- [1] Krumholz HM, Terry SF, Waldstreicher J. Data acquisition, curation, and use for a continuously learning health system. *JAMA* 2016; 316: 1669–1670. DOI: 10.1001/jama.2016.12537
- [2] Touati N, Denis J-L, Roberge D et al. Learning in health care organizations and systems: An alternative approach to knowledge management. *Administration & Society* 2015; 47: 767–801. DOI: 10.1177/0095399712459730
- [3] Berwick DM, Nolan TW, Whittington JC. The triple aim: care, health, and cost. *Health Aff* 2008; 27: 759–769
- [4] Sikka R, Morath JM, Leape L. The Quadruple Aim: care, health, cost and meaning in work. *BMJ Quality & Safety* 2015; 24: 608. DOI: 10.1136/bmjqs-2015-004160
- [5] Bodenheimer T, Sinsky C. From triple to quadruple aim: care of the patient requires care of the provider. *Ann Fam Med* 2014; 12: 573–576. DOI: 10.1370/afm.1713
- [6] Itchhaporia D. The Evolution of the Quintuple Aim: Health Equity, Health Outcomes, and the Economy. *Journal of the American College of Cardiology* 2021; 78: 2262–2264. DOI: 10.1016/j.jacc.2021.10.018
- [7] Heytens H, Walther F, Keßler L et al. [Characteristics of Innovation Fund-supported Intervention Studies: Review and Document Analysis of Study Protocols, Publications and Final Reports]. *Gesundheitswesen* 2021; 83: e20–e37. DOI: 10.1055/a-1448-2412
- [8] Wensing M, Grol R. Methods to identify implementation problems. In: Grol R, Wensing M, Eccles M, Hrsg. *Improving patient care: The implementation of change in clinical practice*. Edinburgh: Elsevier; 2005: 109–120
- [9] Dixon-Woods M, Amalberti R, Goodman S et al. Problems and promises of innovation: why healthcare needs to rethink its love/hate relationship with the new. *BMJ Qual Saf* 2011; 20: i47–i51. DOI: 10.1136/bmjqs.2010.046227
- [10] Wensing M, Grol R, Grimshaw J. Hrsg. *Improving patient care: The implementation of change in health care*. Hoboken, NJ: Wiley-Blackwell; 2020
- [11] May C. A rational model for assessing and evaluating complex interventions in health care. *BMC Health Serv Res*. 2006; 6:
- [12] May C. Agency and implementation: Understanding the embedding of healthcare innovations in practice. Part Special Issue: Place, migration & health 2013; 78: 26–33. DOI: 10.1016/j.socscimed.2012.11.021
- [13] Mercuri M, Baigrie B, Upshur REG. Going from evidence to recommendations: Can GRADE get us there? *J Eval Clin Pract* 2018; 24: 1232–1239. DOI: 10.1111/jep.12857
- [14] Sackett DL. Was ist Evidenz-basierte Medizin und was nicht? *Münch med Wschr* 1997; 139: 644–645
- [15] Katz DL. *Clinical epidemiology and evidence-based medicine: Fundamental principles of clinical reasoning and research*. Thousand Oaks: Sage; 2001
- [16] Schünemann HJ, Guyatt GH. *Clinical Epidemiology and Evidence-Based Health Care*. In: Ahrens W, Pigeot I, Hrsg. *Handbook of epidemiology*. 2. Aufl. New York: Springer Science; 2014: 1813–1873
- [17] Bastian H, Bender R, Ernst AS et al. *Methoden: Version 2.0 vom 19.12.2006*. Köln: Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen; 2006

- [18] Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). Allgemeine Methoden: Version 7.0. 7. Aufl. Köln: Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG); 2023
- [19] Higgins JP, Green S. *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0; 2011
- [20] Higgins JP, Green S. Hrsg. *Cochrane handbook for systematic reviews of interventions*. West Sussex: John Wiley & Sons Ltd; 2008
- [21] Mayntz R, Rosewitz B, Schimank U et al. Hrsg. *Differenzierung und Verselbständigung : zur Entwicklung gesellschaftlicher Teilsysteme*. Frankfurt am Main: Campus; 1988
- [22] Sackett DL, Haynes RB, Guyatt GH et al. *Clinical epidemiology: A basic science for clinical medicine*. Boston. Lippincott: 1991
- [23] Merton RK. The unanticipated consequences of purposive social action. *American Sociological Review* 1936; 1: 894–904
- [24] Clay-Williams R. Complex systems and unintended consequences. In: Rapport F, Clay-Williams R, Braithwaite J, eds. *Implementation science: The key concepts*. Abingdon, Oxon, New York, NY: Routledge; 2022: 193–195
- [25] Mannion R, Braithwaite J. Unintended consequences of performance measurement in healthcare: 20 salutary lessons from the English National Health Service. *Internal Medicine Journal* 2012; 569–574. DOI: 10.1111/j.1445-5994.2012.02766.x
- [26] Magerøy MR, Wiig S. The effect of full-time culture on quality and safety of care – a literature review. *IJHG* 2023; 28: 68–81. DOI: 10.1108/IJHG-11-2022-0106
- [27] Raspe H, Pfaff H, Härter M et al. *Versorgungsforschung in Deutschland: Stand, Perspektiven, Förderung: Stellungnahme ; Standpunkte*. Weinheim: Wiley-VCH; 2010
- [28] Nelke A. Der Innovationsmanagementprozess im Unternehmen. In: Nelke A, Hrsg. *Kommunikation und Nachhaltigkeit im Innovationsmanagement von Unternehmen: Grundlagen für die Praxis*. Wiesbaden: Springer Fachmedien Wiesbaden; 2016: 5–42
- [29] Vahs D, Brem A. *Innovationsmanagement: von der Idee zur erfolgreichen Vermarktung*. Schäffer-Poeschel. 2015;
- [30] Schoffer O, Wimberger P, Gerken M et al. Treatment in Certified Breast Cancer Centers Improves Chances of Survival of Patients with Breast Cancer: Evidence Based on Health Care Data from the WiZen Study. *Geburtsh Frauenheilk* 2024; 84: 153–163. DOI: 10.1055/a-1869-1772
- [31] Schmitt J, Klinkhammer-Schalke M, Bierbaum V et al. Initial Cancer Treatment in Certified Versus Non-Certified Hospitals—Results of the WiZen Comparative Cohort Study. *Dtsch Arztebl Int* 2023 arztebl. m2023.0169. DOI: 10.3238/arztebl.m2023.0169
- [32] Roessler M, Schmitt J, Bobeth C et al. Is treatment in certified cancer centers related to better survival in patients with pancreatic cancer? Evidence from a large German cohort study. *BMC Cancer* 2022; 22: 621. DOI: 10.1186/s12885-022-09731-w
- [33] Völkel V, Gerken M, Kleihues-van Tol K et al. Treatment of Colorectal Cancer in Certified Centers: Results of a Large German Registry Study Focusing on Long-Term Survival. *Cancers (Basel)* 2023; 15:. DOI: 10.3390/cancers15184568
- [34] Bierbaum V, Schmitt J, Klinkhammer-Schalke M et al. Potenzialabschätzung für die Konzentration der Versorgung von Krebspatient:innen in Kliniken mit DKG-Zertifizierung mittels Überlebenszeitanalyse. *Gesundheitswesen* 2023; 85: S197–S204. DOI: 10.1055/a-2132-6797
- [35] Bierbaum V, Bobeth C, Roessler M et al. Treatment in certified cancer centers is related to better survival in patients with colon and rectal cancer: evidence from a large German cohort study. *World J Surg Oncol* 2024; 22: 11. DOI: 10.1186/s12957-023-03262-9
- [36] Eifler S, Leitgöb H. Experiment. In: Baur N, Blasius J, Hrsg. *Handbuch Methoden der empirischen Sozialforschung*. 3. Aufl. Wiesbaden: Springer VS; 2022: 225–241
- [37] Kubbe I. Experimentelle Forschung – neue Entwicklungen. In: Jäckle S, Hrsg. *Neue Trends in den Sozialwissenschaften*. Wiesbaden: Springer Fachmedien Wiesbaden; 2017: 85–114
- [38] Ahmad A, Chowdhury D. Resistance to change during uncertainty-perspective of covid-19: An empirical investigation into resistance management in the healthcare organizations. *REASER* 2021; 22: 5–20. DOI: 10.54609/reaser.v22i2.79
- [39] Shahbaz M, Gao C, Zhai L et al. Investigating the adoption of big data analytics in healthcare: the moderating role of resistance to change. *Journal of Big Data* 2019; 6: 6. DOI: 10.1186/s40537-019-0170-y
- [40] Tretter F, Marcum J. 'Medical Corona Science': Philosophical and systemic issues: Re-thinking medicine? On the epistemology of Corona medicine. *J Eval Clin Pract* 2022. DOI: 10.1111/jep.13734
- [41] Macdonald H, Loder E, Abbasi K. Living systematic reviews at THE BMJ. *BMJ* 2020; 370: m2925. DOI: 10.1136/bmj.m2925
- [42] McPherson SJ, Speed E. NICE rapid guidelines: exploring political influence on guidelines. *BMJ Evidence-Based Medicine* 2021 bmjebm-2020-111635. DOI: 10.1136/bmjebm-2020-111635
- [43] Greenhalgh T. Will COVID-19 be evidence-based medicine's nemesis? *PLoS Med* 2020; 17: e1003266. DOI: 10.1371/journal.pmed.1003266
- [44] Levesque J-F, Sutherland K, Watson DE et al. Learning Systems in Times of Crisis: the Covid-19 Critical Intelligence Unit in New South Wales, Australia. *Catalyst non-issue content* 2020; 1:. DOI: 10.1056/CAT.20.0542
- [45] Pfaff H, Schmitt J. The organic turn: coping with pandemic and non-pandemic challenges by integrating evidence-, theory-, experience-, and context-based knowledge in advising health policy. *Front Public Health* 2021; 9: 1607. DOI: 10.3389/fpubh.2021.727427
- [46] Greenhalgh T, Engebretsen E. The science-policy relationship in times of crisis: An urgent call for a pragmatist turn. *Soc Sci Med* 2022; 306: 115140. DOI: 10.1016/j.socscimed.2022.115140
- [47] Pfaff H, Schmitt J. Reducing uncertainty in evidence-based health policy by integrating empirical and theoretical evidence: An EbM + theory approach. *J Eval Clin Pract* 2023 n/a. DOI: 10.1111/jep.13890
- [48] Hogg HDJ, Al-Zubaidy M, Keane PA et al. Evaluating the translation of implementation science to clinical artificial intelligence: a bibliometric study of qualitative research. *Front Health Serv* 2023; 3: 1161822. DOI: 10.3389/frhs.2023.1161822
- [49] Hughes G, Moore L, Maniopoulos G et al. Theorising the shift to video consulting in the UK during the COVID-19 pandemic: Analysis of a mixed methods study using practice theory. *Part Special Issue: Place, migration & health* 2022; 311: 115368. DOI: 10.1016/j.socscimed.2022.115368
- [50] Subbiah V. The next generation of evidence-based medicine. *Nat Med* 2023; 29: 49–58. DOI: 10.1038/s41591-022-02160-z
- [51] Lee DT, LaCombe J, Chung CK et al. Lag-Time to Publication in Plastic Surgery: Potential Impact on the Timely Practice of Evidence-Based Medicine. *Ann Plast Surg* 2013; 71:
- [52] Lorenz E, Köpke S, Pfaff H et al. Clusterrandomisierte Studien. Teil 25 der Serie zur Bewertung wissenschaftlicher Publikationen (Cluster-randomized studies – part 25 of a series on evaluating scientific publications). *Dtsch Ärztebl* 2018; 115: 163–168. DOI: 10.3238/arztebl.2018.0163
- [53] Hemming K, Lilford R, Girling AJ. Stepped-wedge cluster randomised controlled trials: a generic framework including parallel and multiple-level designs. *Stat Med* 2015; 34: 181–196. DOI: 10.1002/sim.6325
- [54] Lehmann BA, Lindert L, Ohlmeier S et al. "And Then He Got into the Wrong Group": A Qualitative Study Exploring the Effects of Randomization in Recruitment to a Randomized Controlled Trial. *IJERPH* 2020; 17: 1–16. DOI: 10.3390/ijerph17061886

- [55] Busse R, Karagiannidis C, Augurzky B et al. Der Vorschlag der Regierungskommission für eine grundlegende Reform der Krankenhausvergütung. In: Klauber J, Wasem J, Beivers A, Mostert C, Hrsg. Krankenhaus-Report 2023. Berlin, Heidelberg: Springer Berlin Heidelberg; 2023: 267–280
- [56] Sterne JAC, Savović J, Page MJ et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ* 2019; 366: l4898. DOI: 10.1136/bmj.l4898
- [57] Popper KR. *Conjectures and refutations: The growth of scientific knowledge*. 2. Aufl. London: Routledge; 2014
- [58] Rost J. Zeitgeist und Moden empirischer Analysemethoden. In: ZUMA, Hrsg. Forum Qualitative Sozialforschung/Forum: Qualitative Social Research. 2003: 2–3
- [59] Döring N, Bortz J. *Forschungsmethoden und Evaluation*. Wiesbaden: Springer-Verlag; 2016
- [60] Fiedler K, Kutzner F, Krueger JI. The Long Way From α -Error Control to Validity Proper: Problems With a Short-Sighted False-Positive Debate. *Perspect Psychol Sci* 2012; 7: 661–669. DOI: 10.1177/1745691612462587
- [61] Mormina M. Knowledge, Expertise and Science Advice During COVID-19: In Search of Epistemic Justice for the ‘Wicked’ Problems of Post-Normal Times. *Social Epistemology* 2022; 36: 671–685. DOI: 10.1080/02691728.2022.2103750
- [62] Sackett DL, Rosenberg WM, Gray JA et al. Evidence based medicine: What it is and what it isn't. *BMJ* 1996; 312: 71–72
- [63] Stang A. Evaluability of the effect of oncology center certification. *Dtsch Arztebl Int* 2023. DOI: 10.3238/arztebl.m2023.0184
- [64] Schmitt J, Schoffer O, Klinkhammer-Schalke M. Correspondence. *Dtsch Arztebl Int* 2024. DOI: 10.3238/arztebl.m2023.0265
- [65] Roethlisberger FJ, Dickson WJ. *Management and the worker*. An account of a research program conducted by the Western Electric Company, Hawthorne Works, Chicago. Cambridge, Massachusetts: Harvard University Press; 1949
- [66] Mangione-Smith R, Elliott MN, McDonald L et al. An observational study of antibiotic prescribing behavior and the Hawthorne effect. *Health Serv Res* 2002; 37: 1603–1623
- [67] Russo F, Williamson J. Interpreting Causality in the Health Sciences. *International Studies in the Philosophy of Science* 2007; 21: 157–170. DOI: 10.1080/02698590701498084
- [68] Russo F, Williamson J. Epistemic Causality and Evidence-Based Medicine. *History and Philosophy of the Life Sciences* 2011; 33: 563–581
- [69] Greenhalgh T, Fisman D, Cane DJ et al. Adapt or die: how the pandemic made the shift from EBM to EBM+ more urgent. *BMJ Evidence-Based Medicine* 2022; 27: 253. DOI: 10.1136/bmjebm-2022-111952
- [70] Tretter F, Wolkenhauer O, Meyer-Hermann M et al. The Quest for System-Theoretical Medicine in the COVID-19 Era. *Front Med (Lausanne)* 2021; 8: 640974. DOI: 10.3389/fmed.2021.640974
- [71] Watzlawick P, Beavin JH, Jackson DD. *Menschliche Kommunikation: Formen, Störungen, Paradoxien*. Bern: Huber; 2007
- [72] Pfaff H, Zeike S. Hrsg. *Controlling im Betrieblichen Gesundheitsmanagement: Das 7-Schritte-Modell*. Wiesbaden, Heidelberg: Springer Gabler; 2019
- [73] Das deutsche Netzwerk Versorgungsforschung (DNVF). DNVF-Ad-hoc-Kommission publiziert Teil I des Manuals für Methoden und Nutzung versorgungsnaher Daten. *MVF* 2020; 13: II
- [74] Hoffmann F, Kaiser T, Apfelbacher C et al. Versorgungsnaher Daten zur Evaluation von Interventionseffekten: Teil 2 des Manuals. *Gesundheitswesen* 2021; 83: e40. DOI: 10.1055/a-1516-8823
- [75] Tretter F, Peters EMJ, Sturmberg J et al. Perspectives of (I)memorandum for) systems thinking on COVID-19 pandemic and pathology. *J Eval Clin Pract* 2022. DOI: 10.1111/jep.13772
- [76] Sturmberg JP, Martin CM, Katerndahl DA. Systems and Complexity Thinking in the General Practice Literature: An Integrative, Historical Narrative Review. *The Annals of Family Medicine* 2014; 12: 66. DOI: 10.1370/afm.1593
- [77] Paul E, Brown GW, Ridde V. COVID-19: time for paradigm shift in the nexus between local, national and global health. *BMJ Global Health* 2020; 5: e002622. DOI: 10.1136/bmjgh-2020-002622
- [78] Tretter F, Löffler-Stastka H. Medical knowledge integration and “systems medicine”: Needs, ambitions, limitations and options. *Med Hypotheses* 2019; 133: 109386. DOI: 10.1016/j.mehy.2019.109386
- [79] Caniglia G, Luederitz C, von Wirth T et al. A pluralistic and integrated approach to action-oriented knowledge for sustainability. *Nat Sustain* 2021; 4: 93–100. DOI: 10.1038/s41893-020-00616-z
- [80] Banta D. What is technology assessment? *Int J Technol Assess Health Care* 2009; 25: 7–9. DOI: 10.1017/S0266462309090333
- [81] Draborg E, Gyrd-Hansen D, Poulsen PB et al. International comparison of the definition and the practical application of health technology assessment. *International Journal of Technology Assessment in Health Care* 2005; 21: 89–95
- [82] O'Rourke B, Oortwijn W, Schuller T. The new definition of health technology assessment: A milestone in international collaboration. *Int J Technol Assess Health Care* 2020; 36: 187–190. DOI: 10.1017/S0266462320000215
- [83] Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). *Behandlungsgespräche: Führt eine gemeinsame Entscheidungsfindung von Arzt und Patient bei der Therapiewahl zu besseren Ergebnissen?: Vorläufiger HTA-Bericht HT22-01*. Köln; 2023

This article is part of the DNVF supplement “Health Care Research and Implementation”