FORSCHUNGSBERICHTE DES
INSTITUTS FÜR GESUNDHEITSÖKONOMIE UND KLINISCHE
EPIDEMIOLOGIE DER UNIVERSITÄT ZU KÖLN

STUDIEN ZU GESUNDHEIT, MEDIZIN UND GESELLSCHAFT

NR. 3/2009 VOM 26.05.2009 *** ISSN 1862-7412 *** WWW.IGKE.DE/SGMG

Evidence-based medicine and the evaluation of the quality of interventions in prevention and health promotion

Part I: The epistemological chasm can be bridged!

Autoren:

Andreas Gerber, Stephanie Stock, Anna Marie Passon, Markus Lüngen

Korrespondierender Autor: Andreas.Gerber@uk-koeln.de

Die Reihe „Studien zu Gesundheit, Medizin und Gesellschaft“ umfasst Arbeits- und Forschungsberichte des Instituts für Gesundheitsökonomie und Klinische Epidemiologie der Universität zu Köln sowie kooperierender Wissenschaftler und Institutionen.

Die Berichte und weitere Informationen zu den Forschungsberichten können im Volltext abgerufen werden unter
http://www.igke.de/SGMG

Bitte zitieren Sie vorliegenden Bericht als

Gerber A, Stock S, Passon AM, Lüngen M. Evidence-based medicine and the evaluation of the quality of interventions in prevention and health promotion. Part I: The epistemological chasm can be bridged! Studien zu Gesundheit, Medizin und Gesellschaft 2009; Köln: Ausgabe Nr. 3/ 2009 vom 26.05.2009.

Abstract

Introduction: Due to limited resources, interventions in prevention and health promotion must be prioritized using an appropriate assessment instrument. Currently, the three epistemological approaches of best practice, quality assurance and evidence-based evaluation are considered as mutually exclusive. ***Object***: Since the authors planned to devise an assessment instrument to classify studies with regard to their effect, their study quality and the quality of the intervention, a prior epistemological consideration was necessary. **Theory and Methods:** To this purpose, the authors merged the three approaches that have previously been considered mutually exclusive -- (1) best practice; (2) quality assurance, and (3) evidence-based evaluation as all are used in the assessment of prevention and health promotion -- while preserving the strengths of all three. Thereby, the three distinct dimensions can be captured: (1) quality of study design, e.g., prior validation of instruments used to measure effects; (2) quality of intervention, e.g., participation of the intended population, and (3) the outcomes of interventions, e.g., weight loss or time increase in weekly vigorous exercise. ***Results***: As a result of this synthesis of approaches, a criteria catalogue and a model to display the results, the so called "ordinal meta-analysis", could be developed. ***Conclusion***: The currently ideologically dominated debate around methods of evaluation in prevention and health promotion may be resolved and yield an instrument fully accepted by researchers from various backgrounds and well understood by health politicians.

Abstract Deutsch

Einleitung: Auch für Prävention und Gesundheitsförderung gilt das Gesetz der endlichen finanziellen Ressourcen. Daher müssen Interventionen priorisiert werden. Um ein Instrument zur Bewertung von Interventionen zu entwickeln, das eine Priorisierung erlaubt, wird derzeit gerungen. **Theoretische Basis und Methode:** Die Autoren haben drei in der Bewertung von Prävention und Gesundheitsförderung angewandte, aber bisher als sich epistemologisch gegenseitig widersprechend geltende Ansätze unter Bewahrung ihrer jeweiligen Stärken zusammen geführt: 1) Best practice, 2) Qualitätssicherung, und 3) evidenzbasierte Evaluation. Die Stärken und Schwächen der drei Ansätze werden diskutiert. Auf Grundlage des von den Autoren aus der Theorie der Evidenzbasierten Medizin entwickelten Verständnis der evidenzbasierten Evaluation werden Elemente aller drei Ansätze beleuchtet und bewertet. ***Ergebnisse***: Es konnte dargelegt werden, dass die neuartige Synthese der drei Ansätze die drei Dimensionen 1) Qualität des Studiendesigns, z. B. vorausgehende Validierung der Messinstrumente, 2) Qualität der Intervention, z. B. Partizipation der Zielgruppe,

und 3) Ergebnis/ Effekt, z. B. Gewichtsabnahme oder Anstieg der wöchentlichen Zeit in Bewegung, analog zu Prozess-, Struktur- und Ergebnisqualität einschließt. ***Schlussfolgerung***: Auf Grundlage dieser Ergebnisdarstellung können Gesundheitspolitiker/innen Interventionen als empfehlenswert, ablehnenswert oder unsicher, sprich mehr Forschung ist erforderlich, einordnen, wie es der aus dem Ansatz entwickelten Darstellungsform der ordinalen Metaanalyse zu entnehmen ist. Die oft rein ideologische Debatte zwischen den Anhängern der verschiedenen Ansätzen in der Evaluation kann auf Grundlage unserer Überlegungen auf ein neues Niveau gehoben werden.

1 Introduction: The clash in evaluation of prevention and health promotion

Prevention and health promotion and the question of how these can be assessed could easily require separate papers that discuss how each field is defined and understood by various scholars. In this paper, we will first set out brief definitions to establish common ground for the discussion of how to assess interventions in prevention and health promotion. Prevention refers to the avoidance of disease and, thus, is primarily characterized by pathophysiology and patterns and mechanisms of disease genealogy. Not necessarily, but often implicitly, health is perceived as the absence of disease in the context of prevention. On the other hand, health promotion is defined in the Ottawa Charter (1986) as: —…the process of enabling people to increase control over, and to improve, their health. … Health is, therefore, seen as a resource for everyday life, not the objective of living. Health is a positive concept emphasizing social and personal resources, as well as physical capacities. Therefore, health promotion is not just the responsibility of the health sector, but goes beyond healthy life-styles to well-being.[1]

After years of ideological controversy, one may now view prevention and health promotion as different endeavors, but nevertheless, complementary sides of a coin [1]. While the term prevention might be more applicable to immunizations, screening, and programs such as banning vending machines or raising taxes for tobacco, the term health promotion might denote approaches such as multidisciplinary interventions to create exercise-friendly environments or reduce social and economic disparities. Yet, the reduction of risk factors as a core element of prevention and the strengthening of protective factors and resilience as germane to health promotion efforts are often only separable in theory; they are inextricably intertwined in real-world interventions. Hence, we kept the differentiation between prevention and health promotion for heuristic reasons, but integrated the full range of health interventions into our development of a new assessment instrument. But prevention and even more so health promotion face the question: *How do we know what works?*

[1] Ottawa Charter. 1986. http://www.who.int/hpr/NPH/docs/ottawa_charter_hp.pdf.

This is where Evidence-based medicine (EbM) comes into play. Long before the term EbM was coined, as early as 1979, a classification scheme was purposely designed to distinguish tiers of evidence for prevention [2]. Thus, from the very first, the logic of sorting knowledge that is inherent to EbM was applied to prevention. Yet, a rancorous epistemological debate arose over the question as to whether an assessment rooted in EbM could and should be applied to prevention *and* health promotion (for instance [3], [4]). In 2001, the Institute of Medicine (IOM) posited that randomized controlled studies (RCTs) are not applicable as gold standards in the fields of prevention and health promotion. McQueen succinctly summarized: —Few topics in the field of health promotion have engendered as much heated debate as that of evidence. [5].

In this paper, we present the epistemological grounds of our new method that we developed as part of our project to assess interventions from prevention and health promotion. The authors have merged three approaches that have previously been considered mutually exclusive -- (1) best practice; (2) quality assurance, and (3) evidence-based evaluation as all are used in the assessment of prevention and health promotion -- while preserving the strengths of all three. We then developed a criteria catalogue and, subsequently, a new model for synthesizing findings in prevention and health promotion -- our proposed —ordinal meta-analysis that is a tool that displays results of assessments embracing criteria from all three approaches.[2] This model will ensure that assessments meet the needs of the users [6] and allow health policy makers to prioritize interventions in prevention and health promotion using our new instrument that assesses both the quality of study design as well as the quality of intervention while also incorporating the outcomes of interventions. As any scale or assessment instrument needs to be grounded in a theory, in this part of our two-paper series, we will lay epistemological grounds.

[2] The methodology, the criteria catalogue including a manual on how to apply it, and five —ordinal metaanalyses— one of which will be given as an example in this paper—displaying the assessment of more than a hundred interventions in the fields exercising among women and girls, exercising in a workplace environment, nutrition in school-based interventions, prevention of depression in adolescents in school settings, and prevention of smoking in pregnant women are accessible in a book-format publication in German [6].

2 Theorie and Methods: Epistemological Foundation of an assessment instrument

What evidence-based evaluation is and what it isn't

We will maintain that the underlying idea of evidence-based medicine (EbM) can be applied to health promotion and prevention. To this end, we will propose a broader approach to what we will call evidence-based evaluation. We much prefer this term, —evidence-based evaluation, "in health promotion and prevention – and we herein put forth the notion that this moniker is not solely applicable to medicine. By using the term, evaluation, in lieu of medicine, we do not confine health promotion to a mere preventative model of disease avoidance more or less controlled by the medical profession. Neither do we, however, use the phrase,—evidence-based health promotion," as that term might give the impression that prevention merely is an appendix to health promotion.

We begin our discourse on evidence-based evaluation by referring back to a classic definition of evidence-based medicine: "Evidence-based medicine is the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients. The practice of evidence based medicine means integrating individual clinical expertise with the best available external clinical evidence from systematic research." [7] Sackett's „practice of evidence" rests on three columns: (1) external evidence; (2) individual clinical expertise, and (3) patients' choices. Even while we understand that evidence-based medicine is based on these three columns, we will focus first on external evidence, described as the process of „systematically finding, appraising and using contemporaneous research findings ...". [8]. In part, the inclusion of guidelines for practitioners that are based on practitioners' experiences with the implementation of interventions into our new integrative assessment tool may be considered equivalent to the inclusion of clinicians' experiences in EbM according to [7]. Sackett's third element „patients' choices" is equivalent to our incorporation of the intended group's participation and their empowerment.

At present, both the evaluation and rankings of research results are far from being reconciled into one generally-accepted and widely-applicable system. Variations are perceived by advocates of EbM,

for instance [9], and variation is demonstrated in: (1) classification and recommendation schemata; (2) quality scales, and (3) the practice of systematic reviews.

(1) While the Scottish Intercollegiate Guidelines Network (SIGN) [10] four-tiered classification of evidence might appear to be the most widely accepted, there are nevertheless schemata that propose other ways to classify types of research and medical knowledge [11]. Two examples are the Oxford Centre for Evidence-based Medicine Levels of Evidence [12] and the Australian National Health and Medical Research Council's Designation of Levels of Evidence [13], [14], [15].

(2) There are a variety of scales (about 25 in number) for assessing study quality (e.g., [16], [17], [18], [19], [20]). In reference to tools and scores, there are ongoing controversies on establishing standards for: a) how and why items are included, b) the number of items, c) the reliability of scores, and d) the scoring range. Scoring ranges seem to us to be particularly problematic because they employ a single arbitrary threshold between what constitutes a good vs. an unacceptable study design. Results of reviews may vary according to the scale used [21].

(3) Authors of systematic reviews are creative in devising their own lists of quality criteria. The German Institute for Quality and Efficiency in Health Care (IQWiG), the Cochrane Collaboration, and the British National Institute of Clinical Excellence (NICE) all differ in their assessments of study caliber.

In consideration of these three points and their underlying controversies, we demonstrated that EbM is far from establishing a monolithic standard and universally accepted criteria. EbM does not dogmatically favor a certain type of research, but – in our perspective – it solely means forming a hierarchy of scientific results that is constructed in a systematic, transparent, and comprehensible manner based on an initially-posed research question.

Even while some researchers may maintain that the RCT or the systematic review with or without a meta-analysis represents the highest level of evidence, a gold standard, this is true for the SIGN classification, but certainly not for the Oxford scheme. We quite agree that RCTs should be undertaken, whenever appropriate as in drug studies and in many preventative and health promotion trials, however, we the authors advocate the adoption of a classification system that takes into

account the underlying research question. The Oxford classification departs from a rigid one-fits-all system. Depending on the subject of research, therapeutic, diagnostic, or prognostic questions, classifications of study designs differ.

Evidence-based evaluation as consensus process

Rychetnik [22], when speaking of evidence in prevention and health promotion, pointedly state „that the 'concept' of evidence is socially constructed, i.e., what counts as evidence […is] negotiated phenomena." As [23] point out, when scientific data or knowledge are analyzed and synthesized into an evidence-based process, it is nonetheless a consensus process. In any systematic review or in the construction of subsequent guidelines, initially, two or more experts independently assess the quality of studies garnered through systematic research. In doing so, the experts use previously-defined and negotiatedcriteria. Consensus is achieved despite the fact that there are instances of disagreements. In fact, socially-negotiated rankings of categories and outcomes color the scientific research process from ist first stage of constructing study designs to the final review phase.

For illustration, we refer to a recent review [24]. The authors use the words „consensus" and „decision" at *every* stage of collating study results -- at the points of: (1) defining the inclusion criteria; (2) methodological quality assessment, and (3) data extraction. Any consensus and thus any tool or scale embodies some degree of arbitration. [25], for instance, are aware of the dialectic of being systematic and arbitrary at the same time by calling their rating system, „arbitrary", while at the same time, considering it: „a suitable method for distinguishing objectively between different study designs … and to assess the methodological quality of studies … in a reproducible manner. … cut-off points used for categorizing a study as a high-quality study and for concluding consistent results were chosen arbitrarily." [25]. They say, EbM is characterized by reproducibility, transparency, and consensus, and thus it is a product of „rational" mediation and interaction. But, even underlying statistics rely on a consensus process based on stated goals. The explicit goal is to obtain a higher certainty, so alpha or p, for instance, are set preferably low, but alpha/ p <= 0.05 remain figures arrived at by research community agreement.

So EbM *per se* does not deny that any expert position could be as valid as the results of experimental studies, but what the value of EBM is that it points out that such knowledge claims of experts have but only a relative degree of certainty. A single physician's opinion based on their experience, a

methodologically sound prospective experimental study, and a meta-analysis of studies may yield the same knowledge, e.g., that light weight molecular heparine has no advantage over unfractioned heparine for the prevention of thrombosis prophylaxis. But differing degrees of certainty are assigned to various sources of knowledge in EbM. EbM attempts to arrive at a justification and substantiation of potentially conflicting claims to truth, and backs them up with relative degrees of certainty. Yet one must always keep in mind that what counts as sufficient certainty is the result of negotiated categories and socially-constructed values. After some philosophical „reconstruction" of EbM, we now set out how we understand EbM and what aspects of the theory can be retained for the purpose of devising an instrument that informs decision-makers about reliable interventions in prevention and health promotion.

EbM, for us, is not a uniform movement, because it permits both a narrow and a wide perception of what its tenets are. The narrower view of „EbM equals RCT" cannot be sustained. If we go back to the initial definition of EbM as being conscientious, explicit, and judicious, what we have retained is „rational" reproducibility, re-audit, transparency, accountability, and reduction of bias as integral requirements for determining reliability in scientific research. As any research is determined by its initial (clinical) question, the research question should also be taken into account to regulate the resolution of conflicting findings. What remains open to debate is the level of evidence that might be required for non-pharmacological studies. We hold the dialectic view (juxtaposing seemingly-opposite ideas) that RCTs should not be considered the sole highest source of evidence, but that RCTs should nevertheless be undertaken whenever feasible which is more often the case than researchers commonly know [26].

Quality assurance

Putting aside the category of intervention effects, the other two dimensions of quality capture the crucial distinction between using a validated instrument to measure outcomes, for instance, as part of a sound study design[3] and an intervention in which the intended group's participation or the proper training of intermediate professionals are considered „good" or desirable qualities. Assessing the nature of study designs is an integral part of any systematic review as it amplifies the reliability of

[3] In this context, we can only hint at the further distinction between methodological quality vs. quality of reporting, e.g. [8], that we have herein combined in our use of the term, „study quality."

results. Quality of study design refers to its compliance with methodological requirements, such as clarifying the mode of randomization.

Now, why should any researcher group together trials in which professionals were properly trained with those in which professionals were insufficiently trained? Study results might differ and average out to a zero-effect despite the following difference: In one study, intermediate professionals were well trained through instruction and practice for an exercise intervention in schools and in another, professionals only received manuals. Given that both studies were conducted with rigorous study quality, exactly this difference in training could explain the contradictory outcomes and hence should not be „averaged" out. In general, since trials might score very well in caliber of their study designs, but not well in intervention design and vice versa, these two dimensions should be measured on different axes of an assessment scale.[4]

Our tripartite instrument reflects the dimensions of quality developed by [27] whose model is also widely applied to interventions in prevention and health promotion. Donabedian differentiated between the qualities of process, structure, and outcome. In the last, Donabedian's category matches our own category of effect. Donabedian's process quality refers to study design, whereas his term, structure, encompasses many of the same aspects that our use of the term, quality of intervention, would include. Emphasis on study design can be dropped if an intervention goes beyond the phase of trial, but quality of intervention would still be vital for generating an effect. Moreover, both measures of study design and intervention for internal as well as external validity need to be taken equally into account to develop a single synthesizing instrument. Internal validity functions as a barrier against the adoption of false positive results, while external validity warrants transferability and generalization.

Values and assessment of intervention quality

Even though some quality criteria are to various degrees based on a sound statistical methodology calculating correlations between effect/outcome and caliber of study design, e.g. [18], in the end, these criteria are all normative, because each has a pre-determined cut-off point. To make our point

[4] A case will be demonstrated below in Figure 1.

explicit: indicators of quality are „somewhat arbitrary“. But, quality criteria *should be* transparent and explicit, so they can, in fact, satisfy the core requirements of evidence-based evaluation. It is particularly true when assessing the quality of intervention that the dilemma of arbitrary values comes most clearly into the play. If, for instance, patients‘ choices are translated into the context of prevention and health promotion, that freedom of choice could indicate empowerment and participation in an intervention. But what if studies have not yet proven that such choices contribute to better outcomes? The certainty that the results of the study will be reproducible in a new setting might be increased, as the intended group participated in the initial planning of the intervention and thus inhibiting as well as facilitating factors might have been taken into account. If evidence-based practice regards patients‘ choices as one of its core tenets, evidence-based evaluation of health promotion might likewise incorporate democratic values, and include intended group empowerment and participation as an indicator of success.

3 Results

Best practice, quality assurance, and evidence-based evaluation: a synthesis of approaches

At present, there are three approaches or methods used to evaluate interventions in prevention and health promotion. In addition to evidence-based evaluation, as presented here, there are also best practice and quality assurance. All three have advantages and disadvantages with regard to the prioritization of interventions.

The objective of best practice approaches in prevention and health promotion is to define a list of model projects with exceptionally good results that can be adopted by others. These lists advise decision-makers on the selection of interventions and may help to learn from the best. There are, however, several shortcomings: a) the collection of interventions is not systematic (experts include single projects on the list without knowing whether similar projects yielded other effects); b) only studies with positive effects or those that are not yet finished are included (what about studies that had negative results even though performed with high quality and rigor? – these also yield very useful information), and c) the requirements of what constitutes best practice are often very low. The scores for the European project coordinated by the German Federal Centre of Health Information (BZgA) require only three out of twelve criteria [28] to be met. The worthy intention of best practice is not to keep people from launching preventive or health promotion projects, but we must caution that low-bar requirements for best practice cast doubts on the reliability and selection of recommended interventions. Criteria taken from best practice approaches and included in our criteria catalogue, for instance, are requirements for participation and empowerment of intended group.

Quality assurance and management criteria in prevention and health promotion are major factors in the design of instruments that are particularly geared to be used by those who plan and implement interventions [29], [30], [31], [32]. Some of these instruments allow for a score, but they are: a) very lengthy, as their intention is to be exhaustive to underpin the real-world management of interventions, and but these are unsuitable as screening tools, and b) not primarily intended for prioritization but rather for quality assurance in planning interventions for everyday life. Criteria retained from quality assurance and added to our criteria catalogue, for instance, are the nature of

project management, the training of intermediate personnel, and the evaluation of documentation process.

Both approaches seem to overlap in their theoretical bases as well as in their criteria. Both are grounded in the theory of quality assurance and measurement. Moreover, best practice and advisory tools for intervention managers are in part devised by a collegial community of researchers and practitioners. But instruments from both best practice as well as quality assurance unite criteria that are not homogeneous, ergo these cannot be measured equitably nor used in assessment because some of them, such as participation, are value-laden. This is both an advantage, as a fundament for best practice is the Ottawa Charter (see above), but also a disadvantage, as little evidence exists [33] in how far empowerment of the intended population guarantees effectiveness or generates a larger effect through interventions.

Evidence-based medicine/health promotion bases its results on systematic search and the presentation of results in systematic reviews. The advantage is a thorough analysis of all interventions in one field. Yet, the systematic review has its faults. One, only the quality of study design is assessed, but not quality of intervention. Thus, interventions that widely vary with regard to criteria such as participation, empowerment, stakeholder qualification, etc. are not distinguished. These disparities have already been criticized, not only by opponents of EbM, but also by experts who develop new forms of systematic reviews (see above). Moreover, the averaged or synthesized result is stripped of context-sensitive traits and are then lost during the implementation of the recommended intervention in prevention and health promotion. At this point of research into prevention and health promotion, findings from high-ranking studies merit consideration. Criteria from EbM taken up into our catalogue include, for instance, prior power calculation, use of validated measuring instruments, and, if applicable, comparability of control and intervention groups and mode of randomization.

Table 1 clarifies the strengths and weaknesses as well as commonalities and pecularities of the three integrated approaches in order to understand how they could be blended into one instrument.

Table 1: Instruments of Quality Assessment in Prevention and Health Promotion

Instrument	Organisation/ Country	Objective
Closing the Gap [4]	Euro Health Net	List of best practice projects
Kriterien guter Praxis in der Gesundheitsförderung sozial Benachteiligter [5]	Federal Centre for Health Education (BZgA)	Guideline for project management
European Quality Instrument for Health Promotion (EQUIPH) [6]	Getting Evidence into Practice	Improve evaluation of quality of interventions
Interactive Domain Model [7]	Centre For Health Promotion der Universität von Toronto	"Best Practice" Guideline for both project management and decision making
Prevention Effect Management-Instrument (PREFFI) [8]	Nationaal Instituut voor Gezondheidsbevordering en Ziektepreventie (NIGZ)	Quality management and assessment during project management
Quint-Essenz [9]	Swiss Federal Office of Public Health (FPOH) and Stifung „Gesundheitsförderung Schweiz"	Guideline for project management
Qualitätssicherung Prävention (QIP) [10]	Universitätskrankenhaus Hamburg-Eppendorf, Federal Centre for Health Education (BZgA)	Expert assessment of intervention with subsequent expert advice to project management
Koskinen-Ollonqvist [11]	Finland	Prioritization of interventions

This new synthesis of epistemologically distinct approaches in the evaluation of interventions in prevention and health promotion requires a new method to display research findings which we will present here as well.

New modes of synthesizing research findings – feeling one's way beyond the "classical" systematic review

After laying the epistemological ground for our approach, we now turn to the challenge of how to align divergent research findings. The „classical" instrument of EbM is the systematic review in which either the general efficacy or the more context-mediated effectiveness of a number of studies are synthesized into an overall result. That means all study results are pooled and resulting conclusions reflect averages.

A limitation of many systematic reviews is their global focus of inquiry; such overarching views detach findings from contexts. They ask, for instance: *do interventions that change the environment help to increase exercising at work?* But the entire spectrum of interventions and large numbers of intended groups in the systematic review are often too heterogeneous to pool. Yet, on the other hand, inclusion criteria are often so narrowly defined that very few studies remain based on pre-selected parameters as may be seen in [34].

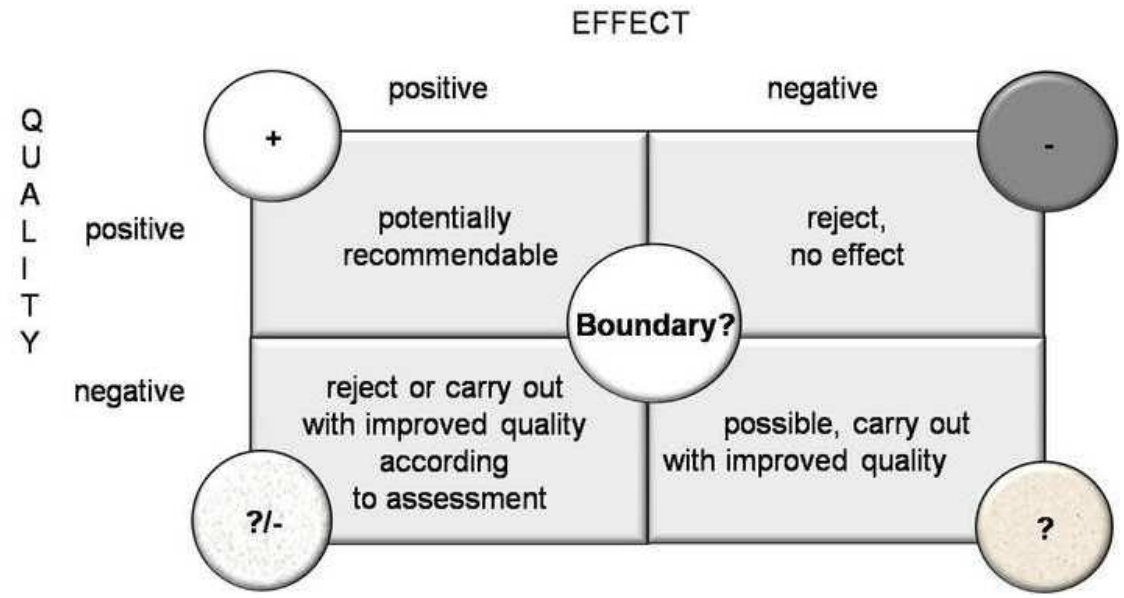
Yet, even if study quality is evaluated (for instance, [35]), the quality of the intervention itself rarely gets considered. Thus, the question -- *why do interventions have positive or negative effects?* – is neglected and remains unanswered (vs. *whether* there are certain effects). Some new approaches merit our attention, particularly publications emanating from the Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-centre), for instance, [36], [37], or from University of Glasgow, for instance, [38], [39]. The EPPI-centre gauges the assessment of individual studies against the factors that would facilitate or inhibit a particular intervention in a manner akin to a cross table.
Thus, their reviews do not aim at an overall pooled effect calculated from a large canon of studies. Of interest to these researchers is not whether an intervention works at all (efficacy: walking reduces coronary heart disease risk), but rather what intervention works for whom in what context and why (effectiveness: increasing walking among 20 to 40 year old mothers in a suburban setting). This approach assures a higher level of external validity in terms of transferability to other settings where similar facilitating and inhibiting factors are encountered.

Like the EPPI-centre, our own aim was not a classical systematic review, and yet we could not follow their pathway. The EPPI-centre, in our view, already delves deeply into an analysis of study

transferability and why some interventions work better than others. Thus, our new method would answer, rather, the preceding question: is the *quality of intervention* in any study good enough for health promotion evaluators to consider a published study a reliable source? Our approach could then be synthesized with that posed by the EPPI-centre, and yield insights for politicians and health managers who must determine which interventions to choose from among several promising recommendations.

So, our question is: how can one separate „good" from „bad" studies and draw up a positive and negative lists of interventions that have a heightened certainty. Ideally, one should be able to assign studies to one of four categories: (1) If a study has an effect and complies with quality criteria, it is more likely that this effect will be replicated when the intervention is implemented elsewhere. (2) Studies of low quality and no effect should not be implemented. The effect might be manifest, but because quality criteria have not been met, results are likely to be false positives. (3) If the study has no effect albeit high scores for quality criteria, it should be put on another list of interventions to be rejected. (4) If a study has a negative effect with poor quality of execution, it should either be redone in compliance with higher standards or it should not be considered for prioritization.

Figure 1



4 Conclusion

The current controversy about which approach should dominate evaluation in prevention and health promotion could be resolved by analyzing evidence-based medicine and building upon its tenets. We, then, could merge evidence-based evaluation with the other two approaches of best practice and quality assurance. Thus, we developed an appraoch building upon the strengths of all three methods than may also facilitate communication among researchers and health politicians as to which interventions to adopt in prevention and health promotion. In addition, we presented an instrument to display results of assessments to decision-makers. In the second paper of this series, we will explain which criteria from all three approaches were included into the assessment tool, the criteria catalogue.

5 Literature

[1] Walter U, Schwartz F.W. Prävention. In:. Schwartz FW, Badura B, Busse R, et al. (Hrsg): Public Health. Gesundheit und Gesundheitswesen. München und Jena: Urban and Fischer, 2003:189-214.

[2] Canadian Task Force on the Periodic Health Exmination. The periodic health examination. Canadian Medical Association journal 1979; 121 (9):1193-1254.

[3] Bödeker W. Evidenzbasierung in Gesundheitsförderung und Prävention – Der Wunsch nach Legitimation und das Problem der Nachweisstrenge. In: Bödeker W, Kreis J. (Hrsg): Evidenzbasierung in Gesundheitsförderung und Prävention. Bremerhaven: BKK Gesundheitsförderung und Selbsthilfe. 2006:12.

[4] Elkeles T. Evaluation von Gesundheitsförderung und Evidenzbasierung? In: Bödeker W, Kreis J. (Hrsg): Evidenzbasierung in Gesundheitsförderung und Prävention. Bremerhaven: BKK Gesundheitsförderung und Selbsthilfe. 2006:111-54.

[5] McQueen D.V. Strengthening the evidence base for health promotion. *Health Promotion International* 2001; 16 (3): 261-268.

[6] Jackson N, Waters E. for the Guidelines for Systematic Reviews in Health Promotion and Public Health Taskforce. Criteria for the systematic review of health promotion and public health interventions. Health Promotion Intern 2005; 20 (4): 367-74.

[7] Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. 1996. Evidence-based medicine: What it is and what it isn´t. British Medical *Journal* 1996; 312 (7023): 71-2.

[8] Rosenberg W, Donald A. Evidence based medicine: An appraisal to clinical problem-solving. British Medical *Journal 1995;* 310 (6987):1122-1126.

[9] Glasziou P, Vandenbroucke J, Chalmers I.Assessing the quality of research. British Medical *Journal* 2004; 328 (7430): 39-41.

[10] Scottish Intercollegiate Guidelines Network (SIGN). No year. Annex B: Key to evidence statements and grades of recommendations.
http://www.sign.ac.uk/guidelines/fulltext/50/annexb.html.

[11] Atkins D, Eccles M, Flottorp S, Guyatt GH, Henry D, Hill S, Liberati A, O´Connell D, Edejer T, Vist GE, Williams HW Jr and the GRADE Working Group. Systems for grading the quality of evidence and the strength of recommendations I: Critical appraisal of existing approaches. BMC Health Services Research 2004; 4: 38.
(http://www.biomedcentral.com/1472-6963/4/38).

[12] Ball C, Sackett D, Phillips B, Straus S, Haynes B. Levels of evidence and grades of recommendations. Last revised 17 September 1998 (http://ww.cebm.net/levels_of_evidence.asp). Centre for Evidence-Based Medicine.

[13] NHMCR (National Health and Medical Research Council). How to use the evidence: assessment and application of scientific evidence. Handbook series on preparing clinical practice guidelines. Reference No CP69. Commonwealth of Australia 2000. Canberra: NHMRC.
(http://www.nhmrc.gov.au/publications/_files/cp69.pdf)

[14] NHMCR (National Health and Medical Research Council). A Guide to the Development, Implementation and Evaluation of Clinical Practice Guidelines. Canberra 1999. NHMRC.

[15] NHMCR (National Health and Medical Research Council). How to review the evidence: systematic identification and review of the scientific literature. Handbook series on preparing clincical practice guidelines. Reference No CP65. Commonwealth of Australia 2000. Canberra. NHMRC (http://www.nhmrc.gov.au/publications/_files/cp65.pdf)

[16] Moher D, Jadad AR, Tugwell P. Assessing the quality of RCT. Current issues and future directions. *International Journal of Technology Assessment in Health Care* 1996; 12 (2):195-208.

[17] Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, McQuay HJ. Assessing the quality of reports of randomized clinical trials: Is blinding necessary? Controlled Clinical Trials 1996 ; 17(1):1-12.

[18] Jüni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. *Journal of the American Medical Association* 1999 ; 282 (11): 1054-60.

[19] Linde K, Clausius N, Ramirez G, Melchart D, Eitel F, Hedges LV, Jonas W.B. Are the clinical effects of homeopathy placebo effects? A meta-analyis of placebo-controlled trials. *Lancet 1997 ;* 350 (9081): 834-43.

[20] Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: An annotated bibliography of scales and checklists. Controlled Clinical Trials 1995; 16(1):62-73.

[21] Jüni P, Egger M. 2003. Qualitätsbeurteilung. In: Antes D, Basler J. (Hrsg) *Evidenzbasierte Medizin.* Stuttgart: Thieme 2003: 60-71.

[22] Rychetnik L, Wise M. Advocating evidence-based health promotion: reflections and a way forward. *Health Promotion International* 2004;19(2):247-57.

[23] Bödeker W, Kreis J. Evidenzbasierte Gesundheitsförderung und Prävention. Methodische Anforderungen und Kenntnisstand. *Gesundheits- und Sozialpolitik* 2003; 57 (9/10): 39-45.

[24] Sluijs EMF van, McMinn AM, Griffin S.J. Effectiveness of interventions to promote physical activity in children and adolescents: systematic review of controlled trials. British Medical *Journal* 2007; 335(7622):703, http://www.bmj.com/cgi/reprint/335/7622/703.

[25] Proper KI, Koning M, van der Beek AJ, Hildebrandt VH, Bosscher RJ, van Mechelen W. Effectiveness of physical activity programs at worksite with respect to work-related outcomes. *Scandinavian Journal of Work Environmental Health* 2002; 28 (2): 75-84.

[26] Rosen L, Orly M, Dan E, Zucker D. In defense of the randomized controlled trial for health promotion research. American Journal of Public Health 2006; 96 (7):1181-86.

[27] Donabedian A. The quality of care. How can it be assessed. *Journal of the American Medical Association* 1988; 260 (12): 1743-48.

[28] Lehmann F, Geene R, Kaba-Schönstein L, Brandes S, Köster M, Kilian H, Steinkühler J, Bartsch G, Linden S. Kriterien guter Praxis in der Gesundheitsförderung bei sozial Benachteiligten. Ansatz Beispiele-Weiterführende Informationen. In: BZgA (Hrsg.) Gesundheitsförderung Konkret Bd. 5, 2006.

[29] Bollars C for the VIG (Flemish Institute for Health Promotion) and the NIGZ (Netherlands Institute for Health Promotion and Disease Prevention) 2005. *User Manual. European Quality Instrument for Health Promotion*. Woerden: NIGZ. http://ws5.evision.nl/systeem3/images/User%20Manual1.pdf#search=%22EQUIHP%20GEP%22.

[30] GEP (Getting Evidence into Practice) in cooperation with NIGZ (Netherlands Institute for Health Promotion and Disease Prevention) und VIG (Flemish Institute for Health Promotion) 2005. *European Quality Instrument for Health Promotion (EQUIHP).* http://ws5.evision.nl/systeem3/images/Annexe%2010%20EQUIHP.pdf#search=%22EQUIHP%20GE P%22.

[31] Kahan B, Goodstadt M. 2005. *THE IDM MANUAL - a guide to the IDM, Centre of Health Promo*tion, http://www.idmbestpractices.ca/idm.php.

[32] Molleman G.R.W. *Health Promotion Effect Management Instrument Preffi 2.0 Assessment Package*. NIGZ 2003 (Netherlands Institute for Health Promotion and Disease Prevention), Woerden, http://www.nigz.nl/index_en.cfm?act=esite.tonen&a=6&b=54&c=69.

[33] Wallerstein N. *What is the evidence on effectiveness of empowerment to improve health?* Copenhagen 2006: WHO Regional Office for Europe (Health Evidence Network report, http://www.euro.who.int/Document/E88086.pdf.

[34] Engbers LH, van Poppel MN, Chin A Paw MJ, van Mechelen W. Worksite health promotion programs with environmental changes. *American Journal of Preventive Medicine* 2005; 29(1):61-70.

[35] Egger M, Jüni P, Bartlett C, Holenstein F, Sterne J. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. *Health Technology Assessmen 2003;t* 7 (1):1-76.

[36] Thomas J, Harden A, Oakley A, Oliver S, Sutcliffe K, Rees R, Brunton G, Kavanagh J. Integrating qualitative research with trials in systematic reviews. British Medical *Journal* 2004; 328 (7446): 1010-1012.

[37] Thomas J, Sutcliffe K, Harden A, Oakley A, Oliver S, Rees R, Brunton G, Kavanagh J. Children and Healthy Eating: A systematic review of barriers and facilitators. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London 2003.

[38] Ogilvie D, Hamilton V, Egan M, Petticrew M. Systematic reviews of health effects of social interventions: 1. Finding the evidence: how far should you go? *Journal of Epidemiology and Community Health* 2005 ; 59 (9): 804-808.

[39] Ogilvie D, Hamilton V, Egan M, Petticrew M. Systematic reviews of health effects of social interventions: 2. Best available evidence: how low should you go? *Journal of Epidemiology and Community Health 2005;* 59(10):886-892.