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

STUDIEN ZU GESUNDHEIT, MEDIZIN UND GESELLSCHAFT

NR. 4/2009 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 II: A criteria catalogue and a new instrument to display results

Autoren:

Andreas Gerber, Evelyn Plamper, Anja Schwalm, Martina Thelen, 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, Plamper E, Schwalm A, Thelen M, Stock S, Passon AM, Lüngen M. Evidence-based medicine and the evaluation of the quality of interventions in prevention and health promotion Part II: A criteria catalogue and a new instrument to display results. Studien zu Gesundheit, Medizin und Gesellschaft 2009; Köln: Ausgabe Nr. 4/ 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. 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. ***Theory and methods*:** The authors propose herein to provide an integrative assessment instrument that is consistent with the core elements of evidence-based medicine as a systematic, explicit, and transparent method of dealing with conflicting research findings. This instrument, a catalogue of criteria, was developed from a synopsis of nine existing instruments to assess interventions in prevention and health promotion through an internal Delphi process. ***Results*:** A criteria catalogue with 68 questions was devised. Among them, seven were ussed to assess the quality of study design as in evidence-based medicine, and twelve were to assess quality of intervention as in best practice and quality assurance. To capture the three distinct dimensions: (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, a new way of synthesizing research was presented, the "ordinal meta-analysis". Health policy makers may now prioritize interventions, recommending some while rejecting others based on their positive or negative effects, using our new integrative assessment tool that ranks the caliber of study designs and the quality of interventions. **Discussion:** Strengths of our criteria catalogue are the transparency and flexibility of the process as well as the way the results are presented. For the time being, the criteria catalogue would need further testing for validity.

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, haben die Autoren 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. ***Theoretische Basis und Methode***: Das daraufhin in einem internen Delphiverfahren auf der Basis von bestehenden international angewendeten Instrumenten, z. B. PREFFI, entwickelte

Bewertungsinstrument von Interventionen in Prävention und Gesundheitsförderung beruht insbesondere auf den Anforderungen der evidenzbasierten Medizin, nämlich systematisch, explizit und transparent mit konfligierenden Studienergebnissen umzugehen. ***Ergebnisse***: Die neuartige Synthese der drei Ansätze hat eine neue Ergebnisdarstellung zur Folge, die sog. „ordinale Meta-Analyse". Die „ordinale Meta-Analyse" schließt 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 ein. Zum Bereich von Interventionen zur Depression bei Adoleszenten wird dies exemplarisch verdeutlicht. Auf Grundlage dieser Ergebnisdarstellung können Gesundheitspolitiker/innen Interventionen als empfehlenswert, ablehnenswert oder unsicher, sprich mehr Forschung ist erforderlich, einordnen. ***Diskussion***: Grundsätzliche Stärke des Kriterienkatalogs und der neuen Form, Ergebnisse zu präsentieren, sind die Transparenz und Flexibilität des Ansatzes. Eine Testung auf Validität des Katalogs ist der nächste Schritt in der wissenschaftlichen Diskussion.

1 Introduction

In 2005, a "Prevention and Health Promotion Act" was passed by the Lower House of the German Parliament (Bundestag), but failed to find the consent of the Federal Council (Bundesrat). Following elections in September of that year, the new government formed by the two largest parties in the House of Parliament, the Christian-Democrats (CDU) and the Social Democrats (SPD), agreed upon a new prevention act in their coalition agreement. So far they failed to draw a new bill. But despite this failure of legislation with a "Prevention and Health Promotion Act" , it is most likely that any intended interventions will have to be assessed prior to implementation. Germany thus needs to establish whether evidence exists to support particular interventions. To this end, we analyzed intervention studies from various countries. On the basis of a certain degree of similarity with regard to culturally held beliefs and attitudes as well as to overall economic performance, studies from 13 countries in Europe, the Americas, and Oceania were selected for analysis. It is also commonly held that these are the countries which have national prevention policies and outstanding projects in prevention and health promotion. Therefore, interventions have to be selected, evaluated and assessed for transferability.

Yet, before this assessment can be performed, an instrument must be developed. Health politicians need reliable and valid assessments of interventions in prevention and health promotion in order to prioritize and build upon effective interventions. In this paper, we present the evolution and applicability 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.[1] This model will ensure that assessments meet the needs of the users [2] and allow

[1] The methodology, the criteria catalogue including a manual on how to apply it, and five "ordinal meta-analyses"—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

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 previously laid out in our paper on the epistemological grounds of our endeavor (part I of this series) to devise a catalogue of criteria, we will explain, while putting aside the category of intervention effects, that the two dimensions of (1) study quality and (2) intervention quality capture the crucial distinction between using a validated instrument to measure outcomes, for instance, as part of a sound study design[2] 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 results. Quality of study design refers to its compliance with methodological requirements, such as clarifying the mode of randomization.

So, our question is: how can one separate "good" from "bad" studies and draw up a positive and negative list of interventions that have a heightened certainty. Ideally, one should be able to assign studies to one of four categories: (a) 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. (b) 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. (c) If the study has no effect albeit high scores for quality criteria, it should be put on another list of interventions to be rejected. (d) 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. Moreover, one must distinguish whether a particular intervention lacks quality because there was either no thorough analysis of the demand or resources in the intended population, or if the fault lies in the poor nature of the study design. Therefore, quality criteria need to be broken into two different dimensions and should not be summarized into a single score.

---

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 [1].

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

2 Methods

In the following section, the development of the evaluation instrument (criteria catalogue) will be described.

The internal Delphi process

Initially, a search was undertaken for instruments to assess interventions in prevention and health promotion. Instruments could be retrieved based on our knowledge, on expert contacts and screening of webpages of organizations dealing with prevention and health promotion. Then, each instrument (see Table 1) was extensively presented to the internal group of six experts (MASG, EP, AS, MT, AG plus one other former colleague). Its objectives, list of criteria, strengths and weaknesses as well as previous application and experiences with the tool were discussed. If the team lacked information, experts representing one of the instruments were contacted directly. A synopsis of the instruments was drawn. Based on the group's understanding of evidence-based evaluation, best practice, and quality assurance – as delineated in part I of this series – an integrated catalogue of criteria was devised through a consensus process. The aim was to devise a feasible catalogue that would hold criteria to discern reliable from less reliable interventions with regard to their effect.

In a second step, the criteria selected for the questionnaire were assigned the categories descriptive and normative as well as quality of study design, quality of intervention, and effect. Criteria to measure quality of study design were based on similar scores used in evidence-based medicine, eg the Jadad-score. Criteria to score quality of intervention were chosen weighing secondary literature, using experience from quality management and assurance and best practice.

The scoring proecedure

As we, the authors, are convinced that studies with various designs, such as RCTs, CTs and even non-CTs should be considered for answering questions in prevention and health promotions (cp. part I of this bipartite series of articles), our scoring had to allow for various maximum levels. Also, with regard to character of campaigns, media or mass campaigns should possibly be included into an assessment with multidisciplinary and participatory interventions in order not to preclude what kind

of campaign might work better. Thus a scoring system was developed that then was transferred into percent of maximum attainable points.

How to display the results?

In contrast to the procedure of "best practice" or a systematic review, we chose to proceed according to what we refer to as "ordinal meta-analysis" approach. Best practice primarily pools studies which are selected without a systematic search, which report an effect and which fulfil selected criteria pertaining to quality of the intervention – criteria pertaining to quality of the study are for the most part not considered – to form a good practice/best practice list. In a systematic review which selects studies on the basis of quality of the study and evaluates the total result in terms of a positive or negative effect, it is not possible to analyse the studies in a differentiated manner. The procedure adopted differs from these two approaches and is depicted in Figure 1. Our approach is a synthesis of the evaluation of quality of intervention found in the best-practice-model and the evaluation of quality of study and effect found in the systematic review. Since it became clear that the classical systematic approach would not be the right method to answer the question of "How an effect of an intervention could be substantiated", the literature was searched for various approaches to display results. Again, our objective was to create an easy understandable and feasible instrument that, however, should integrate the three dimensions of effect, study quality and quality of intervention as independent from each other.

Recommendations

Any collection of data irrespective of caliber and the resulting review needs to be translated into recommnedation. To this purpose, various instruments and devices are in use to help the users rely on recommnedations. Therefore, we also intended to assign the studies to different levels of certainty.

3 Results

The criteria catalogue: Descriptive and normative parts

A criteria catalogue comprising 68 questions was developed on the basis of current quality tools (see Table 1) as well as under consideration of reviews on participation, empowerment and evidence-based evaluation instruments. In addition to study-related criteria, criteria pertaining to the intervention, target group, context and access were compiled. A total of eight questions refer to the quality of the study and twelve questions to the quality of the intervention.

The criteria catalogue consists of criteria that are descriptive or of informational import such as the number of participants as well as quality criteria that then could be calculated into a score. Among the 68 questions (English version attached as appendix) the following twenty criteria were used to score the caliber of study design and intervention. Characteristics of good study designs were: a) clear statement of hypothesis/ goal of study; b) when RCTs were employed, was the mode of randomization explained (point may be deducted, if problems with randomization suggested a confounding of results); c) inclusion of comparable intervention and control groups, if applicable; d) validation of instruments used to measure effects; e) power calculation prior to study; f) intention to treat or per protocol analysis, and g) clear description of statistical methods. Intervention quality was assessed with the following twelve criteria: a) the use of quality instruments to heighten reliability; b) whether evidence was researched prior to initiation; c) evident participation of target group in planning of intervention; d) empowerment planned and measured; e) evidence that intermediate professionals were involved in planning and f) carrying out the intervention; g) intermediate professionals were adequately trained; h) demand for the project was assessed; i) context was taken into account; j) if project management had clear structures; k) if intervention process was documented, and i) where the documentation process was evaluated.

The scoring process

The questions were, for the most part, awarded a zero or one point, although evaluation per intention to treat received one point, per protocol received 0.5 points and the lack of indication as to which of the two modes of statistical evaluation had been applied received zero points. Also, criteria e)-g) above on the caliber of intervention were assigned 0.5 points each in order to not over-

represent the influence of intermediate professionals on the overall score. Maximum scores differed according to the study design applied: RCTs 7, CTs 6, and non-CTs 5 in the section on study quality. Any intervention that would not have required intermediate professionals, e.g., a media campaign, received a lower maximum score in the section on intervention quality. Scores were subsequently translated into percentages for comparability. RCTs and non-CTs could thus attain 100% with 7 or 5 points respectively for study quality.

List of maximum scores (subsequently calculated in percent for comparability)

RCT	7 study quality, 10.5 intervention quality
CT	6 study quality, 10.5 intervention quality
Non-CT	5 study quality, 10.5 intervention quality

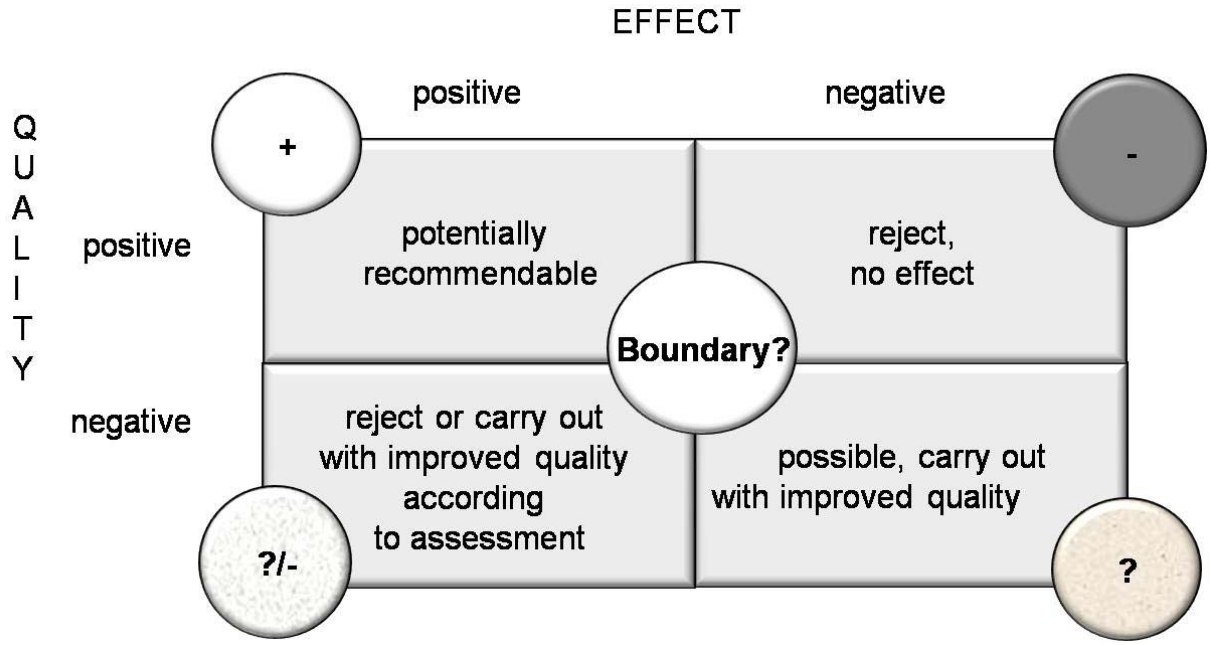
A study without multipliers was awarded 1.5 points less, i.e. maximum score for non-CT media campaign without multipliers: 5 study quality, 9 intervention quality.

Ordinal meta-analysis as a tool to display results to decision makers

In contrast to best practice, quality assurance, or a systematic review, we, the authors, wish to propose a new integrative approach that we call "ordinal meta-analysis" (Figure 1). By the term "ordinal", we refer to a scale that allows for hierarchy (and ordering), but does not assume that the difference between the scores that studies attain translates into a similar difference in their reliability. It only denotes that a higher score means a higher internal or external validity in results. Therefore, an 80% rating vs. 60% is not quantifiably as more valid than a likewise relation of 60% over a 40% rating. "Ordinal" also stands in marked contrast to the standard form of meta-analysis insofar as our conclusions are not the result of mathematical averages of aggregated study results.

Given the great heterogeneity of publications in all fields with regard to intervention and outcome parameters, no mathematical comparison could be made. A kind of "ordinal" meta-analysis was therefore performed, according to which studies were sorted into one of four different classes of recommendation.

Figure 1: "Ordinal meta-analysis" of interventions



Scores for caliber of study design and quality of intervention were calculated independently and then plotted into a two-dimensional plane of axes.

What to recommend?

As scores do not directly translate into recommendations, a four-grade scheme was devised in accordance with similar schemata from evidence-based evaluation [1], as can be seen in Figure 2.

Results of the quality assessment in combination with the assessment of effects yielded the following recommendations:

1) High recommendation for implementation of intervention given positive effect, or non-implementation given negative effect.

Studies with ratings above 80% in both quality of study and intervention quality

2) Strong recommendation for implementation of intervention given positive effect, or non-implementation given negative effect

Studies with ratings above 60% in both quality of study and intervention quality

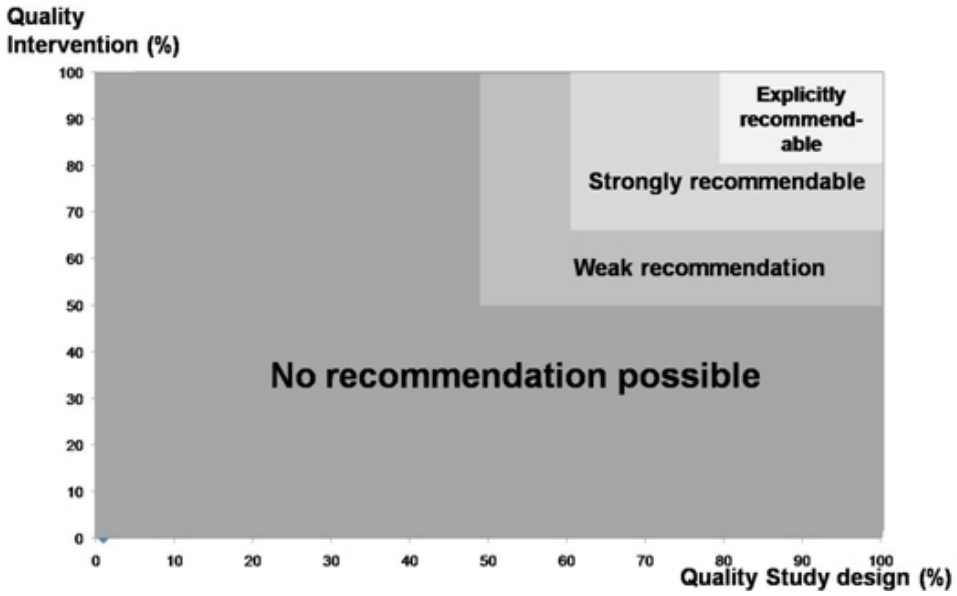
3) Weak (conditional) recommendation for implementation of intervention given positive effect, or non–implementation given negative effect

Studies with ratings above average in each field

4) No recommendation possible

All studies with below average ratings in the assessment of study quality as well as quality of intervention do not permit recommendations as to whether they should be implemented or not.

Figure 2: Recommendations on the basis of scores

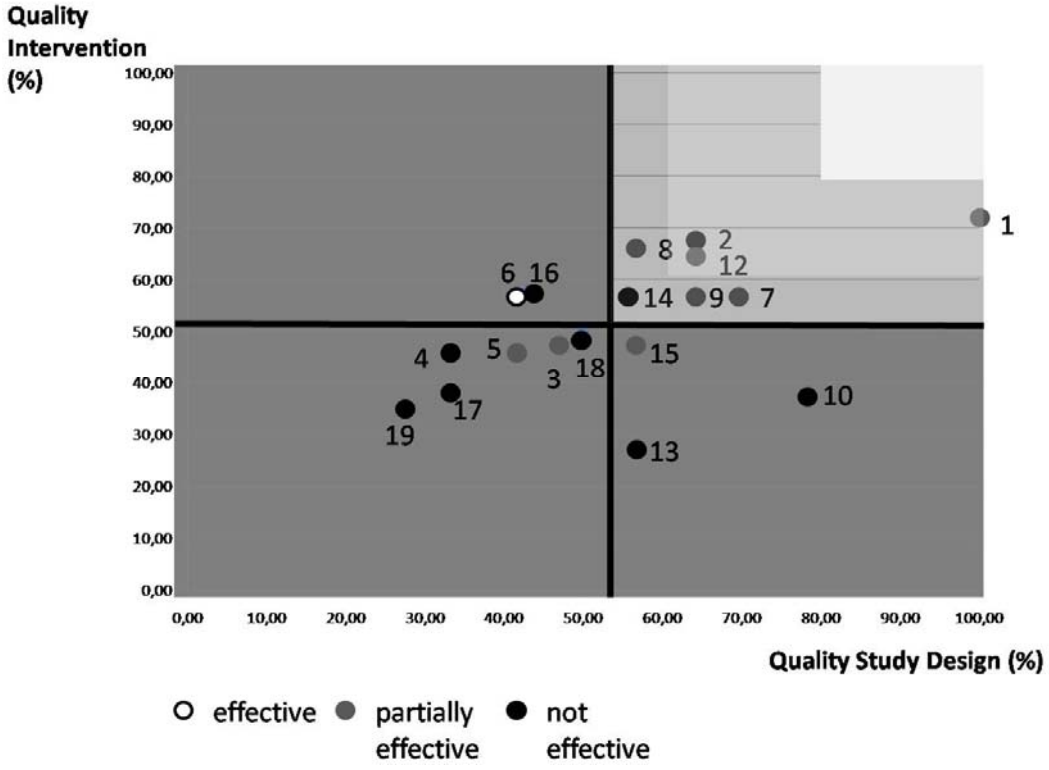


A real case: Interventions to prevent depression in school settings

Results are presented graphically for each of the different intervention areas.

Figure 3 depicts the results of the evaluation of all interventions designed against depression in school settings. Studies are classified according to effect: effective studies are marked as a white dot, non-effective studies as a black dot and partially effective studies, i.e. some positive, some negative results, as a grey dot.

Figure 3: Meta-analysis of interventions against depression in school settings (Adaptation of Figure 13 from [1], p. 112)



Details of the search strategy can be found in [1] (pp. 38-42). Figure 3 displays the results of assessing 18 studies on preventing depression among adolescents in school settings. [12] The horizontal and vertical lines denote averages of study vs. intervention quality respectively. As can be seen, study no. 10 is second best in study quality and would be incorporated into a systematic review by proponents of EbM, whereas we the authors dismissed it because it has a low score in intervention quality. Seven studies ranked above average on both axes and hence were included on positive or negative lists for recommendations (five were partially effective; two were not effective). One international study in the field of depression was rated above 80% in quality of intervention, but none in quality of study design. There was no correlation between effects and quality. No interventions could be classified as highly recommendable, yet three as strongly recommendable [13], [14], [15] and four as of weak recommendation [16], [17], [18], and [19]. The specific strengths and weaknesses of these interventions will be displayed in order to substantiate a recommendation towards implementation (tentative positive list) or non-implementation (tentative negative list). In

advising decision-makers, only these seven studies would be presented indepth so that policies are adopted based on transparent measures of relative success ([1], pp. 113-118).

4 Diskussion

Strengths and weaknesses of the approach

An advantage of our criteria catalogue is that it unites seemingly-contradictory approaches of assessing interventions in prevention and health promotion. Our tool retains the requirements for sound study design espoused by EbM and some of the value-driven criteria such as empowerment and participation of the intended group from best practice and quality assurance. Another strength of our instrument is that scores for caliber of study design and quality of intervention are calculated independently and not summarized into one score that is shaped by a collation of aspects of internal and external validity. The instrument is quite adaptable to various kinds of interventions and to policy-makers' perspectives. Any descriptive category, such as the length of a study or follow-up post-intervention, may receive a score. Thus the ordinal meta-analysis is flexible enough to comply with the wishes or answer questions posed by multiple decision-makers or funders.

Potential problems of our approach should be discussed. Everything is a product of negotiation and consensus. This adage is in accordance with EbM, but it may cause problems, as experts from divergent traditions can devise other criteria catalogues that set different standards for scoring. Yet, as we have noted, this is a recurrent problem in EbM (see Part I). Another problem is the mixture of hard and soft criteria within our catalogue. What seems a rather hard criterion to assess is a prior calculation of power. The training of intermediate professionals, on the other hand, is a softer criterion because there is no clear indicator of how much training is necessary to engender a difference in a study result. Even though we created a lengthy manual ([1], pp. 52-63) for our criteria catalogue, there still was room for debate on whether to give a point or not for the participation of the intended group on the intervention planning process for a particular publication. Again, this is a general problem of EbM that even holds true for the evaluation of the study design quality, a hard criterion, as was demonstrated by [20]). Blind reviewers gave significantly lower scores to study design, which indicates that subjectivity is a factor in *all* scoring. Finally, our criteria catalogue and the

ordinal meta-analysis need more testing with sets of interventions for inter-rater reliability and validity. Yet, we think that one can only compare results derived from various catalogues or instruments and accept them as the consensus of endeavors to be applied to future situations.

5 Conclusion

Our criteria catalogue was devised to separate reliable from non-reliable intervention results. This instrument was based on our understanding of evidence-based evaluation as a process of consensus guided by shared standards and the integration of quality criteria from best practice as well as quality assurance and management in preventative trials [1]. The criteria catalogue permits an integrated assessment of: (1) effectiveness; (2) caliber of study design, and (3) quality of intervention at the same time.

We are not the first ones who have developed a criteria catalogue to evaluate interventions in prevention and health promotion. But our catalogue is unique because it integrates three methods: evidence-based evaluation, best practice, and quality assurance. Thus, our catalogue balances the shortcomings of all three approaches and also integrates practitioners' experiences. It values, moreover, the inclusion of the intended population into planning and enrollment as essential tenets of quality in interventions. As with every new method, this proposed instrument needs thorough testing and evaluation. It is the hope of the authors that this instrument will be applied and tested against heterogeneous data pools and found not only reconcilable but also practical in offering more effective recommendations for health promotion and prevention. Further advantages are that our catalogue of criteria may be adapted to different types of studies as well as different requirements of both intervention and study quality. Cost-benefit analysis (also estimations) may easily be added as criterion.

Finally, we, the authors, strongly recommend that the new "Preventive and Health Promotion Act" should take these results into account with regard to funding decisions that will have to be made. These decisions should be based on a scientific evaluation of evidence, quality of intervention, and quality of study design.

6 Literature

[1] Klever-Deichert G, Gerber A, Schröer MA, Plamper E. International erfolgreiche Interventionen der Prävention und Gesundheitsförderung und ihre Übertragbarkeit auf Deutschland. Studien zu Gesundheit, Medizin und Gesellschaft 2007; Köln: Ausgabe 09/2007 vom 12.11.2007.

[2] 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.

[3] 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.

[4] EuroHealth Net, Bundeszentrale für gesundheitliche Aufklärung. Support Manual „Good Practice", Brüssel: no year.

[5] 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: Bundeszentrale für gesundheitliche Aufklärung (ed.) Gesundheitsförderung Konkret vol. 5, 2006.

[6] 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%20GEP%22., accessed July, 1st, 2007).

[7] Kahan B, Goodstadt M. THE IDM MANUAL - a guide to the IDM, Centre of Health Promotion (http://www.idmbestpractices.ca/idm.php., accessed June 2nd, 2007).

[8] 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., accessed June 2nd, 2007).

[9] Quint-Essenz (http://www.quint-essenz.ch/de, accessed June 2nd, 2007).

[10] Kliche T, Koch U, Lehmann H, Töppich J. Evidenzbasierte Prävention und Gesundheitsförderung. Probleme und Lösungsansätze zur kontinuierlichen Qualitätsverbesserung der Versorgung. Bundesgesundheitsbl- Gesundheitsforsch Gesundheitsschutz 2006;49:141-150.

[11] Koskinen-Ollonqvist P. Developing quality criteria for health promotion programmes in Finland. In: Best Practices. A selection of Papers on Quality an Effectiveness in Health Promotion. Helsinki/ Tallin: May 16-19 1999;205-216.

[12] Lüngen M, Schröer-Günther MA, Passon AM, Gerber A, Lauterbach K.L. Evidenz der Wirksamkeit internationaler Präventionsmaßnahmen und Auswirkungen auf ein deutsches Präventionsgesetz (in press).

[13] Merry S, McDowell H, Wild CJ et al. A randomized placebo-controlled trial of a school-based depression prevention program. J Am Acad Child Adoles Psychiatry 2004;43:538-547.

[14] Pössel P, Baldus C, Horn AB et al. Influence of general self-efficacy on the effects of a school-based universal primary prevention program of depressive symptoms in adolescents: a randomized and controlled follow-up study. Journal of Child Psychology and Psychiatry 2005;46(9):982-94.

[15] Pössel P, Horn AB, Groen G et al. School-based prevention of depressive symptoms in adolescents: A 6-month follow-up. J Am Acad Child Adolesc Psychiatry 2004;43(8):1003-10.

[16] Barrett PM, Farrell LJ, Ollendick TH et al. Long-term outcomes of an Australian universl prevention trial of anxiety and depression symptoms in children and youth: an evaluation of the Friends program. Journal of Clinical Child and Adolescent Psychology 2006;35(3):403-11.

[17] Barrett P, Turner C. Prevention of anxiety symptoms in primary school children: Preliminary results from a universal school-based trial. Br J Clin Psychol 2001;40(4):399-410.

[18] Shochet IM, Dadds MR, Holland D et al. The efficacy of a universal school-based program to prevent adolescent depression. Journal of Clinical Child Psychology 2001;30(3):303-15.

[19] Clarke GN, Hawkins W, Murphy M et al. School-based primary prevention of depressive symptomatology in adolescents. Findings from two studies. J Adolescent Research 1993;8(2):183-204.

[20] 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.