ROBERT KOCH INSTITUT

Short article

Towards a framework for evaluating and grading evidence in public health

Thomas Harder, on behalf of the PRECEPT Expert Meeting Reporting Group[1]

[1] Members of the group are (in alphabetical order):

Muna Abu Sin, Robert Koch Institute, Berlin, Germany

Xavier Bosch-Capblanch, Swiss Tropical and Public Health Institute, Basel, Switzerland

Bruno Coignard, Institut de Veille Sanitaire, Paris, France

Helena de Carvalho Gomes, European Centre for Disease Prevention and Control (ECDC), Stockholm, Sweden

Phillippe Duclos, World Health Organization, Geneva, Switzerland

Tim Eckmanns, Robert Koch Institute, Berlin, Germany

Randy Elder, Centers for Disease Control and Prevention, Atlanta, USA

Simon Ellis, National Institute for Health and Care Excellence (NICE), London, United Kingdom

Frode Forland, Royal Tropical Institute, Amsterdam, The Netherlands

Paul Garner, Liverpool School of Tropical Medicine, Liverpool, United Kingdom

Thomas Harder, Robert Koch Institute, Berlin, Germany

Roberta James, Scottish Intercollegiate Guidelines Network (SIGN), Edinburgh, United Kingdom

Andreas Jansen, European Centre for Disease Prevention and Control (ECDC), Stockholm, Sweden

Gérard Krause, Robert Koch Institute, Berlin, and Helmholtz Centre for Infection Research, Braunschweig, Germany

Daniel Lévy-Bruhl, Institut de Veille Sanitaire, Paris, France

Joerg J. Meerpohl, German Cochrane Center, Department of Medical Biometry and Medical Informatics, University Medical Center Freiburg, Germany

Antony Morgan, National Institute for Health and Care Excellence (NICE), London, United Kingdom

Susan Norris, World Health Organization, Geneva, Switzerland

Eva Rehfuess, Institute of Medical Informatics, Biometry and Epidemiology, University of Munich, Germany

Alex Sánchez-Vivar, Health Protection Scotland (HPS) and the Scottish Health Protection Network (HPN), Glasgow, United Kingdom

Holger Schünemann, Department of Clinical Epidemiology & Biostatistics and Medicine, McMaster University Health Sciences Centre, Hamilton, Ontario, Canada

Anja Takla, Robert Koch Institute, Berlin, Germany

Ole Wichmann, Robert Koch Institute, Berlin, Germany

Walter Zingg, Hôpitaux Universitaires de Genève, Geneva, Switzerland

Teun Zuiderent-Jerak, Institute of Health Policy and Management, Erasmus University, Rotterdam, The Netherlands

Corresponding author: Thomas Harder, MD, MSc; Immunization Unit, Department for Infectious Disease Epidemiology, Robert Koch Institute, Seestrasse 10, 13353 Berlin, Germany. Email: HarderT@rki.de

Abstract

The Project on a Framework for Rating Evidence in Public Health (PRECEPT) is an international collaboration of public health institutes and universities which has been funded by the European Centre for Disease Prevention and Control (ECDC) since 2012. Main objective is to define a framework for evaluating and grading evidence in the field of public health, with particular focus on infectious disease prevention and control. As part of the peer review process, an international expert meeting was held on 13-14 June 2013 in Berlin. Participants were members of the PRECEPT team and selected experts from national public health institutes, World Health Organization (WHO), and academic institutions. The aim of the meeting was to discuss the draft framework and its application to two examples from infectious disease prevention and control. This article introduces the draft PRECEPT framework and reports on the meeting, its structure, most relevant discussions and major conclusions.

1. Introduction

The Project on a Framework for Rating Evidence in Public Health (PRECEPT) has been funded by the European Centre for Disease Prevention and Control (ECDC) since 2012. Main objective is to define a framework for evaluating and grading evidence in the field of infectious disease prevention and control. As part of the peer review process, an international expert meeting was held on 13-14 June 2013 in Berlin. Participants were members of the PRECEPT team and selected experts from national public health institutes, World Health Organization (WHO), and academic institutions. The aim of the meeting was to discuss the draft framework and its application to two examples from infectious disease prevention and control, which were prepared in advance by team members.

PRECEPT was able to build on the work of an ECDC working group, which evaluated the methodology of the Grading of Recommendation, Assessment, Development, and Evaluation (GRADE) Working Group and proposed further discussion of GRADE for application in the context of public health, particularly regarding infectious diseases [1]. It was therefore decided that GRADE will be a key component of PRECEPT, and applying the method to interventional and non-interventional studies is being tested within the project.

According to GRADE, the quality of evidence indicates the extent to which one can be confident that the estimate of effect is correct [2]. GRADE assesses the overall quality of evidence supporting a recommendation across outcomes, which considers the quality rating of all outcomes critical for decision-making. One of four levels of evidence quality is assigned to the review results. Bodies of randomized controlled trials (RCTs) are initially graded as high quality of evidence, whereas bodies of observational studies are initially classified as low quality. Considering a set of criteria might lead to decreasing (downgrading) or increasing (upgrading) one's confidence by one or more levels based on the critical appraisal of the body of evidence related to the outcome under consideration [2].

After an introduction to the PRECEPT framework and the meeting structure, we report the most relevant discussions and major conclusions.

2. The draft PRECEPT framework: overview

The PRECEPT framework – as currently proposed- is intended to rate scientific evidence related to four domains of questions: disease burden, risk factors, diagnostics and interventions. The framework is scheduled into six consecutive steps, from question framing to evidence statement.

In step one, tools are provided to identify key questions relevant for decision-making. Drawing on systematic reviews performed in step two, guidance is provided on the choice of quality appraisal tools (QATs) for assessment of individual studies. An algorithm is given to match a given study design with an appropriate QAT (step three). The set of QATs suggested here has been identified during a review performed by the study team [3]. In step four, a generalized evidence grading based on GRADE is provided to rate the quality of the bodies of evidence. In this step, approaches previously discussed and proposed by the GRADE Working Group [4, 5] or WHO [6, 7] are applied. The latter is used by the WHO Strategic Advisory Group of Experts (SAGE) for the development of vaccination recommendations and includes a modification of the GRADE methodology which allows uprating of evidence quality in the presence of "consistency across investigators, study designs and settings" [7]. For qualitative studies, an approach under discussion by the GRADE Working Group is proposed [8]. The evidence appraisal process ends with the preparation of evidence profiles and summary of findings tables (see [9, 10] for examples) (step five), followed by the preparation of evidence summary statements (step six) (**Figure 1**). By applying this framework, the user should be able to evaluate and grade scientific evidence within the four domains described above in a transparent and reproducible way.

3. Structure of the meeting

Following presentations about the framework and on the application of GRADE to public health, two working groups (WGs) were formed to discuss the draft framework (WG1: from step "framing of questions" to "systematic review"; WG2: from step "quality appraisal" to "evidence summary"), guided by a set of prepared key questions. Participants also split into two WGs to test how different

bodies of evidence from interventional as well as non-interventional studies can be appraised by the framework, using two case studies (WG3 and WG4).

4. Challenges in the application of GRADE to public health

In Randy Elder´s keynote presentation, challenges when applying GRADE to public health were discussed. Two types of challenges were identified. The first one relates to scarcity of evidence from RCTs to address specific public health questions. The second one relates to validity, suggesting that for several public health questions GRADE assessments of evidence quality might be biased and underestimate the true quality of the evidence. For example, studies which measured changes in influenza vaccination coverage attributable to worksite programs [11, 12] received only a "moderate" evidence quality rating according to GRADE. Taking into account factors that are not considered in GRADE ratings, such as the consistently observed step-function increase in vaccination coverage when these programs were implemented, a "high" quality rating would better reflect the true risk that the apparent intervention effects were spurious. Additional examples for community-based interventions were presented, where, according to the speaker, GRADE ratings might not adequately reflect the quality of the evidence derived from specific types of non-randomized studies, such as interrupted time series performed in different settings by different investigators during different time periods demonstrating similar effects [13]. It was therefore proposed that non-randomized designs which are less prone to bias such as a body of interrupted time series should already initially be judged in the GRADE system as being of "moderate" quality. Furthermore, the question was raised whether upgrading of evidence quality might be appropriate when the overall pattern of evidence across settings or study designs mitigates otherwise plausible threats to validity.

5. Summary of discussions: From framing the questions to systematic reviews

The WG suggested revising the framework's domains, including incorporation of additional domains (e.g., "preferences and values", "cost-effectiveness"). Regarding question framing, it was proposed to

use the PI(E)CO framework (P- Population, I – intervention or E- exposure, C – comparator, O – outcome) as standard, and to integrate other questions – especially for domains not related to interventions – as far as possible. Furthermore, the development of a logic model or "conceptual framework", that visualizes the interconnectedness between e.g. risk factors, disease, hospitalization rate and intervention, can be useful to identify relevant questions (for details, see [14]).

The WG proposed that qualitative studies, relevant not only because of their importance for decision-making processes, should not be handled as a separate domain (as suggested in the initial framework proposal), but where appropriate should be integrated in the domains as one of several research designs. Regarding the applicability of GRADE to qualitative studies, the WG concluded that some of the GRADE criteria appear applicable (i.e., indirectness). However, this remains unclear for other GRADE criteria (i.e., imprecision), leaving an area for methods development currently explored by the working group.

6. Summary of discussions: From quality appraisal to evidence summary

For quality appraisal of individual studies, PRECEPT provides a selection of QATs according to study design. The WG concluded that a flowchart is useful to identify the appropriate QAT(s) for a given study design. However, caution is needed when different QATs are used because some are checklists whereas others produce overall quality ratings. It is therefore necessary to interpret the result of the quality appraisal process as a considered judgment, rather than a score, when assigning a quality label to an individual study.

Regarding the applicability of GRADE to domains other than interventions, the group concluded that its application is in principle possible, and that GRADE criteria for downgrading could be applied to all four domains.

For the domains "risk factors" and "interventions" some participants suggested introducing an additional upgrading criterion "consistency across settings and study designs" [6]. Even though intuitively this criterion looks similar to the downgrading criterion "inconsistency", a number of

7

participants found this to be a useful and important criterion that embraces the concept of complementary evidence, which entails more than just the opposite of inconsistency. That is, whereas in the GRADE approach "inconsistency" refers to heterogeneity of effect measures for a given outcome [15], "consistency across settings and study designs" means that consistent results have been obtained under a variety of conditions (study design and setting), which might increase confidence in the overall assessment unless it has to be assumed that an unmeasured confounder influenced the results in all settings in a similar way.

7. Case study I: Introduction of routine rotavirus vaccination of infants in a population

Case study I focused on the question whether routine rotavirus vaccination for infants should be introduced in a specific country and was based on a recently developed recommendation for Germany [10]. The aim was to test the applicability of PRECEPT to grade the available evidence related to local disease burden (or baseline risk) and disease perception. The WG suggested that the particular QAT must be able to account for problems of data quality in different study designs, which are inherent to active vs. passive surveillance systems. Underestimation (as well as overestimation) of disease rates in surveillance systems can be conceptualized as risk of bias and could be estimated through specific studies (e.g. capture-recapture studies). Observational evidence from both active and passive surveillance would enter the GRADEing procedure initially as high quality and should then be downgraded if e.g. a risk of bias is identified. In respect to the systematic literature search it was highlighted by the group, that especially for incidence studies there is often evidence from non-peer-reviewed, "grey-literature" sources such as national disease reporting systems that needs to be identified through specific search strategies.

The WG concluded that in principle it is possible to assess bodies of evidence on disease incidence/prevalence, utilizing the GRADE approach [4]. With regard to the GRADE criterion of "inconsistency", differences in incidence/prevalence data might be explainable by regional differences or temporal trends. It should then be discussed whether it is appropriate to downgrade

evidence quality. In addition, incidence estimates might differ inherently between studies due to their design (active vs. passive case ascertainment). In such situations, one would rather present separate results for the different regions/times/designs.

The WG also concluded that publication bias is important with regard to prevalence/incidence studies because of factors such as pharmaceutical industry interests, in addition to publication practices. WG participants suggested that publication bias in prevalence/incidence studies might be detected by stratification of findings by source of funding.

8. Case study II: Spread of carbapenemase-producing Enterobacteriaceae

Case study II dealt with the spread of carbapenemase-producing Enterobacteriaceae and was based on an ECDC Technical Report [16]. The aim was to test the applicability of PRECEPT to questions related to risk factors and complex interventions. The WG suggested that risk factor studies might in general start as "high quality" of evidence. Thereafter, GRADE criteria would be applied to the respective body of evidence for downgrading, but details remain to be worked out in close collaboration with the GRADE Working Group.

Regarding complex interventions, it is not sufficient to appraise the quality of a body of evidence on a complex intervention with multiple optional components as such. Rather, details on the intervention components must be provided (e.g. in a matrix).

9. Conclusions

The following points summarize the most important meeting results:

- Overall, the participating experts found the proposed PRECEPT framework very helpful and an important step forward in the development of a framework for evidence grading in the area of infectious disease prevention and control.

- It was highlighted that PRECEPT should not be considered as an alternative to GRADE, but rather as a comprehensive framework to help provide guidance in infectious disease
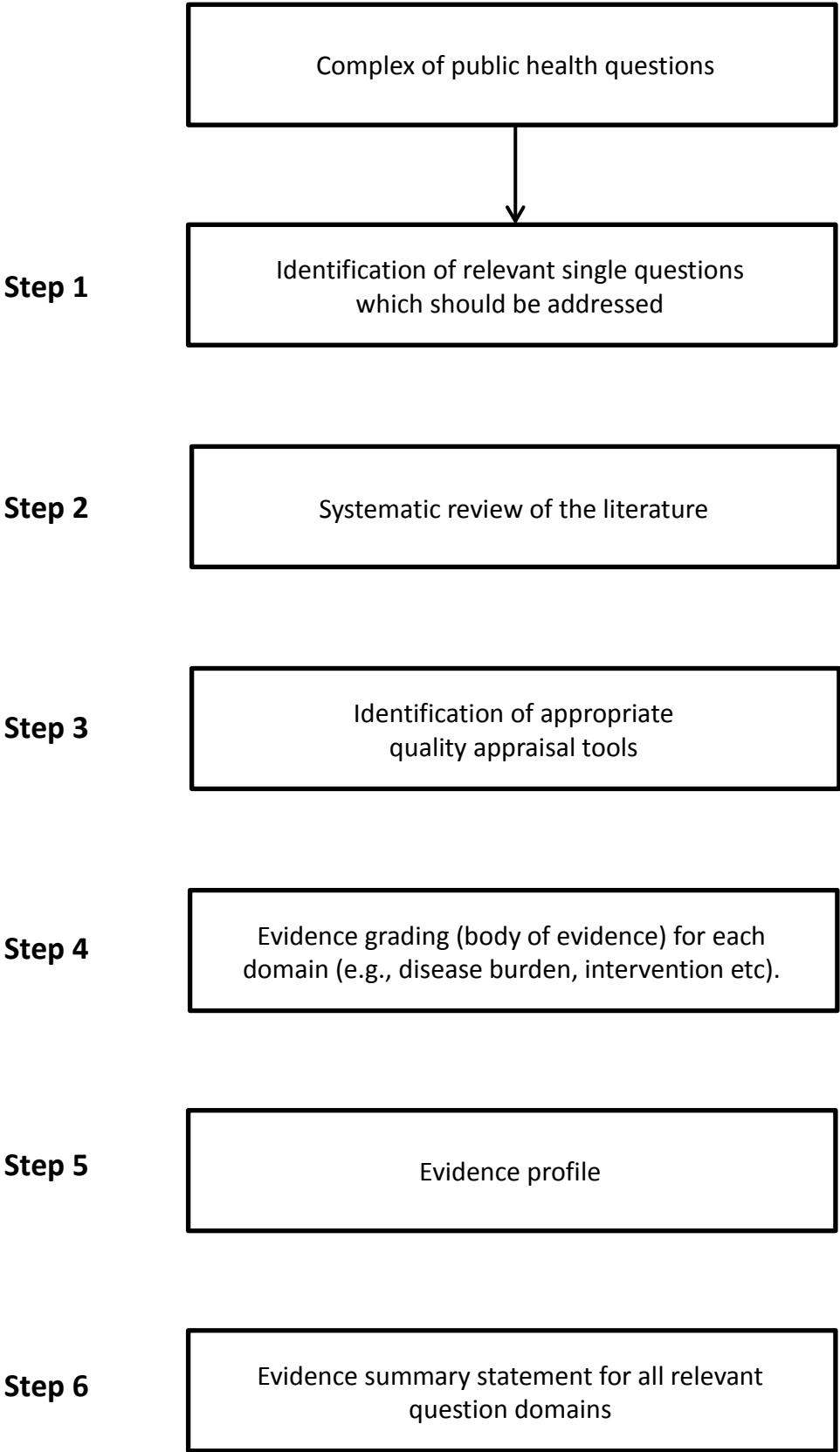
prevention and control decision-making, in which GRADE is integrated as an important component.

- Identifying and formulating the correct questions is the first and most important step in an evidence rating framework. The PI(E)CO approach should be the standard for framing of questions, and questions not related to interventions should be integrated as far as possible. Identification of relevant questions can be supported by construction of a logic model or "conceptual framework".

- Appropriate QATs should be used to assess risk of bias in individual studies, but the results of the appraisal process should be interpreted as a considered judgment.

- GRADE can be successfully applied to bodies of evidence on incidence/prevalence, based on the approach published by the GRADE Working Group.

- Risk factor studies should enter the GRADE-system initially as "high quality" of evidence. Thereafter, GRADE criteria for downgrading and upgrading should be applied, but details remain to be worked out in collaboration with the GRADE working group.

- For the application of GRADE to the domains "risk factors" and "interventions", meeting participants suggested to further discuss whether upgrading of the evidence quality should be allowed for "consistency across settings and study designs".

- GRADE can be successfully applied to evidence on complex interventions; details on the intervention components should be provided.

- Qualitative studies are important at various stages, but should not be handled as a separate domain but as a study design to address given research questions e.g. related to values and preferences. Their quality should be appraised by QATs in the same manner as quantitative studies. Further research is needed regarding the applicability of the GRADE methodology to bodies of evidence from qualitative studies.

Acknowledgments

Figure 1: Flow chart of the PRECEPT methodology

	Complex of public health questions
Step 1	Identification of relevant single questions which should be addressed
Step 2	Systematic review of the literature
Step 3	Identification of appropriate quality appraisal tools
Step 4	Evidence grading (body of evidence) for each domain (e.g., disease burden, intervention etc).
Step 5	Evidence profile
Step 6	Evidence summary statement for all relevant question domains

References

[1]   Evidence-based methodologies for public health - How to assess the best available evidence when time is limited and there is a lack of sound evidence Stockholm: European Centre for Disease Prevention and Control (ECDC), 2011.

[2]   Balshem H, Helfand M, Schunemann HJ, Oxman AD, Kunz R, Brozek J, Vist GE, Falck-Ytter Y, Meerpohl J, Norris S, Guyatt GH. GRADE guidelines: 3. Rating the quality of evidence. Journal of clinical epidemiology 2011; 64:401-6.

[3]   Harder T, Takla A, Rehfuess E, Sanchez-Vivar A, Matysiak-Klose D, Eckmanns T, Krause G, de Carvalho Gomes H, Jansen A, Ellis S, Forland F, James R, Meerpohl JJ, Morgan A, Schunemann H, Zuiderent-Jerak T, Wichmann O. Evidence-based decision-making in infectious diseases epidemiology, prevention and control: matching research questions to study designs and quality appraisal tools. BMC medical research methodology 2014; 14:69.

[4]   Spencer FA, Iorio A, You J, Murad MH, Schunemann HJ, Vandvik PO, Crowther MA, Pottie K, Lang ES, Meerpohl JJ, Falck-Ytter Y, Alonso-Coello P, Guyatt GH. Uncertainties in baseline risk estimates and confidence in treatment effects. Bmj 2012; 345:e7401.

[5]   Schunemann HJ, Oxman AD, Brozek J, Glasziou P, Jaeschke R, Vist GE, Williams JW, Jr., Kunz R, Craig J, Montori VM, Bossuyt P, Guyatt GH, Group GW. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. Bmj 2008; 336:1106-10.

[6]   Duclos P, Durrheim DN, Reingold AL, Bhutta ZA, Vannice K, Rees H. Developing evidence-based immunization recommendations and GRADE. Vaccine 2012; 31:12-9.

[7]   WHO SAGoESo. Guidance for the development of evidence-based vaccine-related recommendations 2014.

[8]   Bohren MA, Hunter EC, Munthe-Kaas HM, Souza JP, Vogel JP, Gulmezoglu AM. Facilitators and barriers to facility-based delivery in low- and middle-income countries: a qualitative evidence synthesis. Reproductive health 2014; 11:71.

[9]   Kredo T, Ford N, Adeniyi FB, Garner P. Decentralising HIV treatment in lower- and middle-income countries. The Cochrane database of systematic reviews 2013; 6:CD009987.

[10]   Koch J, Wiese-Posselt M, Remschmidt C, Wichmann O, Bertelsmann H, Garbe E, Hengel H, Meerpohl JJ, Mas Marques A, Oppermann H. Background paper to the recommendation for routine rotavirus vaccination of infants in Germany. Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz 2013; 56:957-84.

[11]   Lopes MH, Sartori AM, Mascheretti M, Chaves TS, Andreoli RM, Basso M, Barone AA. Intervention to increase influenza vaccination rates among healthcare workers in a tertiary teaching hospital in Brazil *. Infection control and hospital epidemiology : the official journal of the Society of Hospital Epidemiologists of America 2008; 29:285-6.

[12]   Sartor C, Tissot-Dupont H, Zandotti C, Martin F, Roques P, Drancourt M. Use of a mobile cart influenza program for vaccination of hospital employees. Infection control and hospital epidemiology : the official journal of the Society of Hospital Epidemiologists of America 2004; 25:918-22.

[13]   Campbell CA, Hahn RA, Elder R, Brewer R, Chattopadhyay S, Fielding J, Naimi TS, Toomey T, Lawrence B, Middleton JC, Task Force on Community Preventive S. The effectiveness of limiting alcohol outlet density as a means of reducing excessive alcohol consumption and alcohol-related harms. American journal of preventive medicine 2009; 37:556-69.

[14]   Joffe M, Mindell J. Complex causal process diagrams for analyzing the health impacts of policy interventions. American journal of public health 2006; 96:473-9.

[15]   Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, Alonso-Coello P, Glasziou P, Jaeschke R, Akl EA, Norris S, Vist G, Dahm P, Shukla VK, Higgins J, Falck-Ytter Y, Schunemann HJ, Group GW. GRADE guidelines: 7. Rating the quality of evidence--inconsistency. Journal of clinical epidemiology 2011; 64:1294-302.

[16]   Risk assessment on the spread of carbapenemase-producing Enterobacteriaceae (CPE). Stockholm: European Centre for Disease Prevention and Control (ECDC), 2011.