ecdc
EUROPEAN CENTRE FOR DISEASE PREVENTION AND CONTROL

TECHNICAL REPORT

Evidence-based methodologies for public health

www.ecdc.europa.eu

ECDC TECHNICAL REPORT

Evidence-based methodologies for public health

How to assess the best available evidence when time is limited and there is lack of sound evidence

Suggested citation: European Centre for Disease Prevention and Control. Evidence-based methodologies for public health – How to assess the best available evidence when time is limited and there is lack of sound evidence. Stockholm: ECDC; 2011.

Stockholm, September 2011

ISBN 978-92-9193-311-2

doi 10.2900/58229

Foreword

The target audience for this document are public health professionals and policymakers who work in the field of collecting, analysing and evaluating scientific evidence for the purpose of giving evidence-based public health advice for communicable diseases. By this report, we want to address the questions of giving advice under uncertainties, in complex situations and often on short notice.

In the founding regulation of the European Centre for Disease Prevention and Control (ECDC), it is stated that ECDC was established to enhance the capacity of the European Community and the Member States to protect human health through the prevention and control of human diseases – to identify, assess and communicate current and emerging threats by communicable diseases. According to Article 3 in Regulation 851/2004, ECDC shall; (a) search for, collect, collate, evaluate and disseminate relevant scientific and technical data; (b) provide scientific opinions and scientific and technical assistance, including training; and (c) provide timely information to the Commission, the Member States, Community Agencies and international organisations active within the field of public health.

Working in the field of infectious disease epidemiology brings some specific challenges; the case of an infection is at the same time a patient to be cared for and a possible source of spread to others. A case is also an exposure. The questions of incubation periods, transmission routes, infectivity and risks of spread between people can only be studied while an outbreak is ongoing, and rarely experimentally. This poses some specific challenges to the methods of collecting and analysing data from the infectious diseases field.

By this report, our purpose was to explore how the best from the methods of classical epidemiology can be matched and possibly merged with the methodologies developed in evidence-based medicine (EBM) to give a new blend, an evidence-based methodology for infectious diseases prevention and control.

In this document you will find a general background chapter at the beginning of the report that focuses on some of the main topics and challenges of working with evidence-based methods in a public health setting. Each chapter also contains a small background and methods section to make it easier for those who want to read only parts of the chapters independently.

In the process of developing this report, we went through a combination of methods: doing literature searches, consulting experts, facilitating international cooperation between institutions and experts of different specialities, consensus building among experts and carrying out external hearings in the Advisory Forum and in Member States.

There is a wealth of literature on evidence-based methods, and it is impossible to summarise and review them all in a short report like this one. However, for anyone who wants to delve deeper into any aspect, there are comprehensive reference lists at the end of each chapter.

Thank you to those who participated.

Johan Giesecke
Chief Scientist

Table of contents

Abbreviations

AGREE	Appraisal of Guidelines Research and Evaluation
AHRQ	US Agency for Healthcare Research and Quality
CDC	Centres for Disease Control and Prevention (US)
CEBM	Centre for Evidence Based Medicine
COMPUS	Canadian Optimal Prescribing and Utilization Service
CONSORT	Consolidated Standards of Reporting Trials
EBM	Evidence-based medicine
EBPH	Evidence-based public health
ECDC	European Centre for Disease Prevention and Control
EU	European Union
GRADE	Grading of Recommendations Assessment, Development and Evaluation Working Group
HICPAC	Healthcare Infection Control Practices Advisory Committee (US)
HTA	Health Technology Assessment
NHMRC	National Health and Medical Research Council (Australia)
NICE	National Institute for Health and Clinical Excellence
ORION	Outbreak Reports and Intervention Studies of Nosocomial Infection
RCT	Randomised controlled trial
SBAR	Situation, Background, Assessment, Recommendation
SIGN	Scottish Intercollegiate Guidelines Network
STROBE	Strengthening the Reporting of Observational Studies in Epidemiology
TREND	Transparent Reporting of Evaluations of Nonrandomised Designs
USPSTF	US Preventive Services Task Force
USTFCPS	US Task Force on Community Preventive Services
WHO	World Health Organization

Executive summary

The objective of this report is to explore how methods of evidence-based medicine (EBM) can be applied in public health in the field of infectious diseases.

A working group was established with members from EBM organisations, public health institutions and from ECDC, to address questions related to giving evidence-based advice in situations where there is little evidence and shortness of time, to evaluate the need for adaptation of EBM tools for a public health setting, to assess the usefulness of existing guideline development tools and the use of consensus methods in a setting of evidence-based public health. Finally, future challenges and research needs should be identified.

Evidence-based public health could be defined as integrating the best available evidence with the knowledge and considered judgements from stakeholders and experts to benefit the needs of a population. Data from observational studies, surveillance and modelling play an important role as evidence base in public health in the field of infectious diseases. Since information about outbreaks can only be gathered while an outbreak is ongoing, there is a need to better perform and report outbreak investigations. Uncertainties can arise at all stages of a public health decision-making process, or while producing a risk assessment. It is important to handle uncertainties explicitly and transparently and to communicate them to the policymakers. As time goes by and access to evidence increases, uncertainties can be reduced.

A five-stage framework for rapid risk assessments is presented, which includes a preparatory phase and further stages of risk detection/verification, assessment of the risk, development of the advice and implementation and evaluation. Practical tools and templates for each stage are also presented, and the importance of being prepared and having tools at hand when an outbreak occurs is underlined.

The usefulness of evidence-based methods and grading tools are explored. A variety of methods for reporting, assessing and grading evidence are identified and the applicability of these tools in a public health setting is discussed. Special attention has been given to the GRADE instrument, which has been adopted by many influential organisations. A table with a list of issues concerning the applicability of GRADE in public health is presented in Annex 5. Many tools required to produce evidence-based advice already exist, but there is a need to further develop instruments and checklists for some of the study designs relevant to public health.

The AGREE II instrument was developed to evaluate the quality of guidelines. An evaluation of this tool for the purpose of infectious diseases guidelines is presented. The importance of the different domains and items is discussed and some additional criteria for communicable diseases guidelines are proposed. A shorter and adapted version of the AGREE II instrument (GET 5), to be used in time-limited situations, is presented in Annex 7. Further developments and research should focus on how recommendations are influenced by considerations other than the evidence from published literature and on tools for evaluating the professional content of a guideline.

Consensus methods can be used both to evaluate the evidence and to improve the balance of subjective interpretation of evidence from systematic reviews and also to develop best available expert judgements in settings with lack of evidence. Consensus methods can be applied by members of a guidance development group and as a method to facilitate implementation in hearing processes among stakeholders. The importance of transparency in public health decision-making, the role of experts and, finally, how to apply different consensus methods under different timelines are discussed.

By exploring how EBM methods could be applied in public health advice under different time scales, we have found that many methods, tools and templates are already developed, well suited for public health needs and should be more widely used. In some areas there is a need to further develop and fine-tune some the instruments to better fit the needs of public health in the area of infectious diseases. Prioritised areas for future work should be:

- to develop better templates for reporting and checklists for assessing the quality of outbreak reports;
- to explore how evidence-based grading systems could integrate and possibly upgrade the value of evidence from observational research in situations where only such evidence is attainable;
- to develop better retrieval and search systems for observational studies to find the best available evidence under time constraints; and
- to assess how to more explicitly express uncertainties from the scientific evidence and the considered judgements to better inform decision-makers when a public health guidance is given.

The mandate

i. The group should develop methods and tools to strengthen evidence-based work within the public health/infectious disease field – with special attention to situations where time is limited and there is lack of sound evidence, or where methodological standards are less well formalised.

ii. The group should try and adjust EBM tools, standards and checklists on how to collate and evaluate evidence and data to better support public health infectious disease control guidance. In such cases, the best available evidence is often limited to sources like: surveys, studies of outbreaks, studies of infectivity, studies of sero-epidemiology, contact tracing, ecological studies, modelling of epidemics and studies of the natural history of diseases. If need be, proposals for new standards and checklists should be developed. Special attention should be given to situations where time is limited.

iii. The group should evaluate the quality and usefulness of existing grading systems and tools for grading evidence and recommendations for the field of public health/infectious diseases, and suggest adaptations if necessary.

iv. The group should further assess how standards like the AGREE instrument and 'Guidelines for guidelines', primarily developed for the clinical field, can be adjusted and made useful for evaluation and development of public health guidance documents in the field of infectious diseases.

v. The group should consider strengths and weaknesses of different consensus methods.

vi. The group should assess and propose further research needs in the field of methods development in the area of public health/communicable diseases.

vii. A final report should be written and the main conclusions from the working group should be published in a peer-reviewed medical journal.

The members of the working group

EBM institutions

Roberta James, Scottish Intercollegiate Guidelines Network (SIGN), Scotland, UK;
Signe Flottorp, Norwegian Knowledge Centre for the Health Services and GRADE working group, Norway;
Antony Morgan, National Institute for Health and Clinical Excellence (NICE), England, UK;
Simon Ellis, National Institute for Health and Clinical Excellence (NICE), England, UK;
Françoise Hamers, Haute Autorité de Santé (HAS), France;
Jacek Mrukowicz, Institute for Evidence Based Medicine, The Jagiellonian University Krakow, Poland and the GRADE working group (participated in two meetings).

Public health institutions

Alex Sanchez-Vivar, Health Protection Scotland, Scotland, UK;
Anders Tegnell, Swedish Board of Health, Sweden (Member of ECDC Advisory Forum);
Gérard Krause, Robert Koch Institute, Germany (Member of ECDC Advisory Forum);
Heather Murdoch, Health Protection Scotland, Scotland UK;
Hanne Nøkleby, Norwegian Institute of Public Health, Norway (Member of ECDC Advisory Forum);
Lukas Murajda, Jessenius Faculty of Medicine, Comenius University, Slovakia;
Stephen Palmer, Health Protection Agency, UK and Cardiff University, Wales, UK;
Pentti Huovinen, National Institute for Health and Welfare, Finland (participated in one meeting).

ECDC

Helena de Carvalho Gomes, Scientific Advice Unit;
Piotr Kramarz, Scientific Advice Unit;
Paulo Moreira, Health Communication Unit, replaced by Ines Steffens, Health Communication Unit;
Andrew Amato, Surveillance Unit;
Howard Needham, Scientific Advice Unit;
Katrin Leitmeyer, Preparedness and Response Unit;
Ana-Belén Escriva, Scientific Advice Unit;
Marc Struelens, Scientific Advice Unit;
Andreas Jansen, Scientific Advice Unit;
Frode Forland, chair, Scientific Advice Unit.

The big group split into four working groups responsible for each part of the mandate:

- Group A, led by Stephen Palmer, mandate point ii)
- Group B, led by Signe Flottorp, mandate point iii)
- Group C, led by Alex Sanchez-Vivar, mandate point iv)
- Group D, led by Marc Struelens, mandate point v)

1 Background

For the European Centre for Disease Prevention and Control (ECDC) and the European Union Member States' public health institutes to be Centres of Scientific Excellence in the Public Health Field it is of the greatest importance to work according to the best standards of evidence collection, appraisal and application. A more widespread use of evidence-based strategies in public health will foster accessibility of higher-quality information on outcomes of public health interventions, and enhance success of public health policies [1].

A group of experts from 12 countries with a broad experience in different aspects of public health, methodology and infectious diseases were appointed to give guidance to ECDC on how to strengthen its scientific work by adapting and applying methods from evidence-based medicine (EBM) in the field of public health. The group was formed to bring EBM institutions and public health institutions together, as well as the different units within ECDC. This diversity aimed at ensuring that tools developed are both methodologically sound, but also practically based and applicable within the infectious diseases/public health environment.

There are several important distinctions between evidence-based strategies in public health and in clinical medicine, which impede a simple transition of methodologies from one field to the other. When compared to available evidence in clinical medicine, for example, public health evidence is generally a result of the more complex interaction of the best available evidence from research and other epidemiological sources with judgements of the needs, recourses, local circumstances, and ethical and legal implications. Furthermore, there is a need to more adequately assess the scientific quality and risk of bias in the types of studies that form the core evidence base for communicable disease prevention and control, such as outbreak investigations and evidence from observational studies. Finally, evidence-based approaches need to be supported in situations where uncertainties might be high and the time to produce advice is limited.

The influenza A(H1N1) pandemic in 2009/10 clearly demonstrated the need for public health institutes and for ECDC to provide advice in a setting of limited time and lack of sound evidence. For ECDC, there was a need to give quick and considered advice, to weigh benefits and potential harms to Member States and the EU Commission, to support and sometimes coordinate Member States when some very costly decisions had to be taken. The need to share knowledge as it emerges, to execute joint actions and to show solidarity between Member States is continuously being evaluated in such a situation.

It is, therefore, vital that public health institutes and ECDC build processes to ensure that evidence-based approaches are developed for use in both crisis-driven and more reflective scenarios.

- EBM is usually dated from Archibald Cochrane, who in 1972 published a landmark book, 'Effectiveness and Efficiency'. In 1992 the Cochrane Collaboration was established with an aim of making up-to-date, accurate information about the effects of healthcare readily available worldwide. Building up a library of systematic reviews of research evidence has been its greatest achievement [2].
- Working 'evidence based' means applying the scientific principles of transparency, validity and reproducibility when assessing and evaluating the evidence. The process of evidence-based medicine is often referred to as a five stage process: 1) defining the health problem; 2) searching for evidence; 3) assessing the quality of the evidence; 4) implementing the evidence; and 5) monitoring and evaluation.

In a clinical setting, EBM was the application of epidemiological methods to clinical questions, and it became a critical movement among students and clinicians who traditionally had learnt their curriculum by reading authoritative textbooks and by listening to professors sharing experiences and judgements. Concepts like numbers needed to treat, relative and absolute risk reduction, meta-analyses and strengths and weaknesses of different research designs were introduced in the clinical fields of diagnosis, studies of treatment effects, studies of harm and prognosis. The types of question should decide which study design to use to avoid bias, e.g. for intervention studies the classical evidence hierarchy was established with systematic reviews of randomised controlled trials on top, followed by randomised controlled trials, cohort studies and case control studies.

1.1 Evidence-based public health

Evidence-based public health (EBPH) has lagged behind the EBM movement and only relatively recently have the methods started to be applied to the more complex public health problems; and there has been even less spread of these methods into the fields of infectious disease epidemiology. Within the last few years, it has become increasingly obvious that there is a clear need to merge these different cultures and to strengthen public health decisions, using the methods primarily developed by epidemiologists, which are now further developed into tools, networks and processes by the evidence-based medicine movement. In Britain, the National Institute for Clinical Excellence (NICE) was set up to publish EBM guidelines for clinical practice in 1999 and joined with the Health Development Agency in April 2005 to become the new National Institute for Health and Clinical Excellence (still

abbreviated as NICE). The public health section was set up in April 2005 and published its first guidance in March 2006.

EBPH deals with two main applications of evidence [1]. The first application, which is most clearly associated with the term evidence-based medicine, is the evidence of effectiveness of interventions and their translation into programmes and policies. The second area of application of EBM in public health has to do with defining the health risk, identifying groups at special risk, elucidating causal pathways, aetiology, preventable risk factors and assessing the impacts of disease spread and the benefits of prevention.

1.2 Evidence for effectiveness of interventions

Public health evidence for the effectiveness of interventions is generally weak. Thacker et al. [3], at the Centres for Disease Control and Prevention (CDC – US), in 2005, reviewed the evidence for the effectiveness of interventions to modify 194 potentially modifiable risk factors for 31 conditions of high priority for CDC. Of the 702 population-based interventions, evidence for the preventable fractions were found for only 4.4%.

Another example of this issue was illustrated by the pandemic flu 'Technical Report – Guide to public health measures', published by ECDC [4]. An assessment was made of the type and strength of evidence for each of 27 interventions. The grading of evidence was:

Grade A: Systematic reviews where there are diverse primary studies to draw from (not primarily modelling), well-designed epidemiologic studies or especially experimental studies (randomised controlled trials).

Grade B: evidence based on well-designed epidemiologic studies, substantial observational studies or experimental studies with 5 to 50 subjects, or experimental studies with other limitations (not having influenza as an endpoint, for example).

Grade C: evidence based on case reports, small poorly controlled observational studies, poorly substantiated larger studies, application of knowledge of mode of transmission, infectiousness period etc.

Only 2 of 27 – the vaccine studies – were grade A and 13 were grade C, thus emphasising that evidence for the effectiveness of public health interventions often does not meet the criteria used in evidence-based medicine because of the study design used in the evidence-based public health.

Evidence-based medicine approaches are probably most easily applied when dealing with pharmaceutical or other clinical interventions for well-characterised diseases in well-defined populations of patients. The CONSORT guidelines, which stands for Consolidated Standards of Reporting Trials [5], were first published in 1996 and was a necessary step in order to build an evidence base by combining information from individual trials. In public health practice there are many examples of interventions that have been evaluated by randomised controlled trials (RCTs), such as vaccines or population screening and for which the evidence base is usually dealt with in an exemplary way. It should be noted, however, that there are also examples of trials of social policy such as free school breakfasts that have been evaluated by RCTs, although many may have thought it to be impossible. One challenge for the field of public health is to persuade policymakers of the possibility and desirability of trialling interventions.

There are many settings in which RCTs are impossible either because of ethical and logistical issues or because the very rarity of the diseases precludes trials. Trials are particularly difficult to envisage in emerging disease situations both because of the rarity of the first cases, the speed with which response is required and the unacceptability by policymakers and by the general public of having either control groups or different interventions in different parts of the same population. It also has to be acknowledged that there are weaknesses of the EBM approach even when dealing with RCT evidence. Kemm [6] has argued that these weaknesses are even starker in public health. Context is all important in interpreting and applying RCT results. What works in the trial of a carefully defined and narrow group of patients with a specific condition cannot be applied outside those constraints without clinical judgment playing a major part. In public health every population may be considered unique because of local economic, social and cultural factors, thereby challenging the applicability of evidence from any other population, and emphasising the role of judgment.

With the exception of vaccines and antibiotic treatments, the evidence used in communicable disease control is most likely to be derived from microbiological and virological principles coupled with observational studies.

Much of communicable disease control guidance comes from surveillance studies and outbreak investigations. Inroads have been made in developing quality measures for surveillance systems (e.g. nosocomial infection and vaccine surveillance) and some countries have undertaken rigorous evaluation (e.g. Scotland). Good surveillance data has been used successfully to assess the effectiveness of control programmes such as the initial success of the *Hemophilus pneumoniae* vaccine programmes and the subsequent need to introduce a booster dose.

Outbreak data can also be used to evaluate interventions, although doubts have been expressed such as the comments of the authors of the ORION Statement [7]: 'Because many important biases may be in operation and

several interventions are often made simultaneously, outbreak reports are of limited value for assessing the effectiveness of interventions'. Nevertheless in many acute and rare incidents information from previous similar or analogous outbreaks is all the epidemiology there is to go on. Voirin et al. [8] have applied the ORION method to reviewing nosocomial outbreaks of influenza. They showed the great value of a standardised and rigorous approach to describing and summarising characteristics of outbreaks, and give reasons for optimism that risk factors for transmission and the effectiveness of interventions may be illuminated by evidence synthesis from outbreaks if improved investigation and reporting can be achieved. Heijne et al. [9] have offered an innovative approach to evaluating the effectiveness of hygiene measures to control norovirus outbreaks in summer camps using the rate of reduction of the reproduction number.

1.3 EBM methods applied to characterising threats and risks

The second area of application of EBM in public health has to do with defining the health risks. Here the 'best available evidence' may include studies of microbiology and virology, analytical epidemiological studies, descriptive investigative epidemiology, population surveillance and population mathematical modelling.

These methods are sometimes characterised as 'low quality evidence' in the EBM hierarchy, but as with 'higher' forms of evidence the quality of such studies, their collation and interpretation should be guided by EBM methods. This means the application of rigorous, standardised and systematic ways of handling evidence so that the risk of bias is minimised and assumptions are made explicit. As with synthesising intervention data from RCTs and quasi-experimental studies, a rigorous systematic approach to conducting and reporting observational studies is vital if they are to be compared and information collated. The STROBE guidelines for strengthening the reporting of observational studies, especially cohort and case control studies, have been developed to improve the quality of publications in this field, and the MOOSE guidance has been produced for the meta-analysis and systematic reviewing of observational studies [10].

The STROBE working group have argued that 'Incomplete and inadequate reporting of research hampers the assessment of the strengths and weaknesses of the studies reported in the medical literature. Readers need to know what was planned (and what was not), what was done, what was found, and what the results mean. Recommendations on the reporting of studies that are endorsed by leading medical journals can improve the quality of reporting' [11]. The guidelines set out specific criteria for reporting different types of epidemiological studies.

The interpretation of observational studies has special pitfalls especially when causal inferences are wanted. In the UK, the Academy of Medical Sciences has published guidelines on how to identify the environmental causes of disease from non-experimental studies [12]. With regard to findings from case studies with relatively strong causal claims the report says that,

> *'First, they either concerned a very large effect (as with smoking and lung cancer) or they applied to rare and unusual outcomes with distinctive features (as with the fetal alcohol syndrome or the sequelae of profound institutional deprivation or neural tube defects or vCJD). Second, detailed careful attention was paid to alternative non causal explanations and to how to test for their possible role. Third, all made use of multiple research designs (including 'natural experiments') with complementary strengths and limitations. Thus, the smoking research included the study of reversal effects, as did the study of institutional deprivation. Furthermore, adoption and twin designs were used to check the possibility of genetic mediation (as with abuse of children). Fourth, the causal inference was tested in multiple populations that differed in their characteristics. Fifth, animal models and human experimental studies contributed support on biological processes (as with smoking, fetal alcohol syndrome, and the sequelae of institutional deprivation, folic acid and HIV). It is also the case that the apparent success stories stand out in terms of the rigour of both their measurement and their statistical analyses. In no instances did one design provide the 'clinching' proof but, in combination, they made the causal inference a compelling probability.'*

1.4 Uncertainties at different stages of public health

The EBM culture has made some great efforts and improvements in displaying uncertainties and variabilities in the core scientific evidence of a decision-making process. Less has been done up to now to acknowledge the uncertainties of the stages before and after dealing with the core evidence content.

Environmental health sciences and toxicology can be seen as another arm of public health dealing with exposure measurement, new substances and their potential effects on human health and the environment. Expression and communication of uncertainties is a main issue in these fields and infectious disease/public health and the evidence-based medicine field could learn from these developments. The Codex Alimentarius Working Principles for Risk Analysis state: 'Constraints, uncertainties and assumptions having an impact on the risk assessment should be explicitly considered at each step in the risk assessment and documented in a transparent manner. Expression of

uncertainty or variability in risk estimates may be qualitative or quantitative, but should be quantified to the extent that is scientifically achievable'[13].

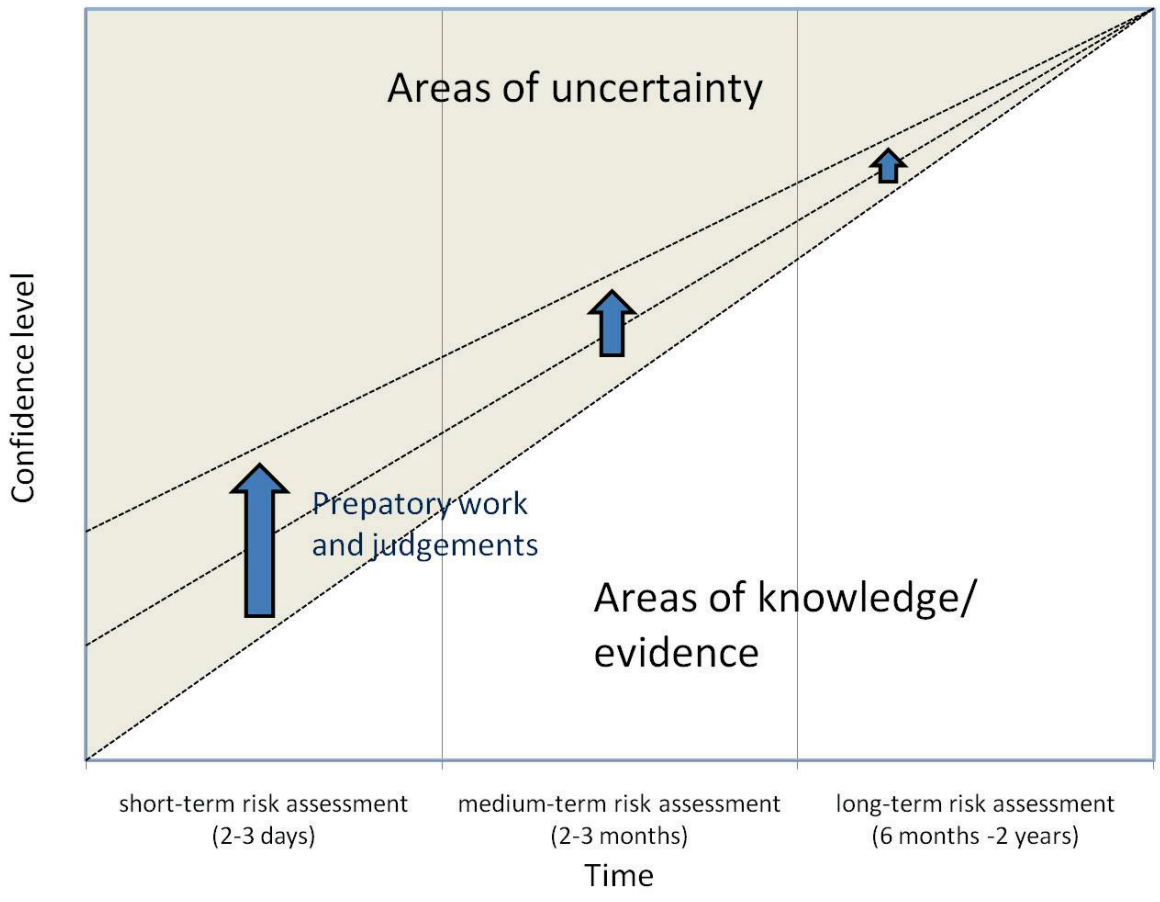
Thus there are three basic requirements for addressing uncertainties in risk assessments: 1) systematically identify and evaluate the sources of uncertainties; 2) evaluate their combined effect on the outcome of the assessment; and 3) communicate this to the risk managers (policymakers).

The EU Scientific Committee for New and Emerging Health Risks (SCHENIR) uses the expression 'lines of evidence' to characterise different sources and levels of evidence and information. They consider lines of evidence that normally are considered to be in the area at the bottom of the EBM hierarchy. We see, however, that this evidence not fitting well into the traditional EMB evidence pyramid and is applicable to other types of questions or other fields than the clinical medical setting. Examples of such lines are studies of exposure, lab research, animal experiments and mathematic modelling experiments. These are study designs relevant to public health in many instances. Application of the precautionary principle is also relevant to judgements of evidence in a public health setting; no evidence of harm should not be interpreted as if there is evidence for no harm.

A system to grade or quantify uncertainty in public health decision-making should be able to handle all stages of a public health decision-making process, and, as well, to incorporate data from the lines of evidence besides the excising evidence hierarchy.

When doing a rapid risk assessment, there will normally be a higher level of uncertainty than in settings where there is more time to gather and evaluate evidence and to involve more experts and stakeholders in the judgements (see illustration below).

Figure 1 Relation between uncertainty and time



The figure illustrates the relation between level of confidence and time. The confidence level increases over time, while areas of uncertainty decrease as more information is being collected and analysed. The relevance of preparatory work, e.g. readily available up-to-date information and optimised work processes and judgments from experts and stakeholders, is more important for short-term (rapid) risk assessments. The confidence level that can be achieved for short-term risk assessments is largely dependent upon the preparatory work done. When there is more time and more resources available, more rigorous methods of collecting, assessing and judging the evidence

can be applied. This figure also implies that the applicability and relevance of the traditional EBM methods increase as time increases. But the principles of EBM, working transparently and according to best available evidence, apply at all timelines.

Decision-making under uncertainty is part of public health. To deal with uncertainties explicitly is necessary to inform policymakers and the public honestly. Uncertainties can arise at all steps and stages of a public health guidance development process. It is an aim to reduce uncertainties to a minimum, but even in a well-performed and strictly reported Cochrane systematic review there are several decisions taken by the researchers where subjective judgement are done. Public health decision-making is often complex, stepwise and multifaceted [14–15].

The more factors that are involved in the decision-making process, the more potential uncertainties arise and the less time there is, the more challenging it is to deal with it.

1.5 References

[1] Brownson RC, Fielding JE, Maylahn CM. Evidence-based public health: a fundamental concept for public health practice. Annu Rev Public Health. 2009 Apr 29;30:175-201.

[2] The Cochrane Collaboration. Available from: http://www.cochrane.org/.

[3] Thacker SB, Ikeda RM, Gieseker KE, Mendelsohn AB, Saydah SH, Curry CW, et al. The evidence base for public health informing policy at the Centers for Disease Control and Prevention. Am J Prev Med. 2005 Oct;29(3):227-33.

[4] European Centre for Disease Prevention and Control. Guide to public health measures to reduce the impact of influenza pandemics in Europe: 'The ECDC Menu'. Stockholm: ECDC; 2009. Available from: http://ecdc.europa.eu/en/publications/publications/0906_ter_public_health_measures_for_influenza_pandemics.pdf.

[5] CONSORT. Welcome to the CONSORT Statement Website. Available from: http://www.consort-statement.org/.

[6] Kemm J. The limitations of 'evidence-based' public health. J Eval Clin Pract. 2006 Jun;12(3):319-24.

[7] Stone SP, Cooper BS, Kibbler CC, Cookson BD, Roberts JA, Medley GF, et al. The ORION statement: guidelines for transparent reporting of outbreak reports and intervention studies of nosocomial infection. J Antimicrob Chemother. 2007 May;59(5):833-40.

[8] Voirin N, Barret B, Metzger MH, Vanhems P. Hospital-acquired influenza: a synthesis using the Outbreak Reports and Intervention Studies of Nosocomial Infection (ORION) statement. J Hosp Infect. 2009 Jan;71(1):1-14.

[9] Heijne JC, Teunis P, Morroy G, Wijkmans C, Oostveen S, Duizer E, et al. Enhanced hygiene measures and norovirus transmission during an outbreak. Emerg Infect Dis [serial on the Internet]. 2009; 15(1): Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2660689/pdf/08-0299_finalR.pdf.

[10] Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA 2000 Apr 19;283(15):2008-12.

[11] STROBE Statement (website). Aim and use of STROBE. Available from: http://www.strobe-statement.org/index.php?id=strobe-aims

[12] Identifying the environmental causes of disease: how should we decide what to believe and when to take action? London: Academy of Medical Sciences; 2007.

[13] Joint FAO/WHO Food Standards Programme. Codex Alimentarius Commission: procedural manual. Section III. 15th ed. Rome: FAO; WHO; 2005. Available from: ftp://ftp.fao.org/codex/Publications/ProcManuals/Manual_15e.pdf.

[14] Stirling A. Keep it complex. Nature. 2010 Dec 23;468(7327):1029-31.

[15] Joffe M, Mindell J. Complex causal process diagrams for analyzing the health impacts of policy interventions. Am J Public Health. 2006 Mar;96(3):473-9.

2 How to give evidence-based guidance when evidence is scarce and the time is limited?

2.1 Aims and objectives

In order to support a public health response to any outbreak of communicable disease there has to be an alerting system with rapid identification of a health threat or potential cause of a disease based on the best available evidence against a background of rumor and 'noise'. Once a risk has been verified, a risk assessment must follow, which is informed by more evidence and expert opinion. Within ECDC this process is defined as a rapid risk assessment (timeline 2–3 days).

A rapid risk assessment should ideally encompass identification of the hazard, the population at risk, existing/implemented control measures, and evaluation of risk, monitoring, updating and documentation. For a rapid risk assessment to be useful, it should include documented gaps in the evidence at each stage with assumptions made. The risk assessment aims to provide advice and guidance for actions, e.g. control measures to be implemented and based on a combination of available evidence (existing, evolving or extrapolated) complemented by expert opinion.

The focus of this chapter is within the stage of a rapid risk assessment and production of advice and guidance. The aim was to assess current evidence-based (EB) approaches, consider how they can be applied and – if necessary – develop a methodology that can support the specific needs of ECDC and other public health institutions that provide public health scientific advice to prevent and control communicable disease.

Two significant challenges that were not fully handled by current EB systems were explicitly addressed. Firstly, there is a need to better assess scientific quality in the types of studies that form the core evidence base for communicable disease prevention and control, which tend to be mainly observational studies, e.g. case control or cohort studies at best. Within the field of evidence-based guidance for public health there is rarely evidence from the traditional higher end of the evidence hierarchy, e.g. randomised control trials. Secondly, there is a need to support evidence-based approaches in situations where time to produce outputs is limited, whereas standard evidence-based methods of guidance production can take from six months to two years.

A more detailed operational methodology for rapid risk assessments has been developed in parallel on a commissioned basis by the Health Protection Agency, UK, and the Evidence-based Methodology Working Group has reviewed the document and had a consultative meeting with the contractor. An executive summary of this tool is to be found in the annex of this document.

2.2 Background

The benefit of a systematic evidence-based approach to the development of guidance in clinical medicine has long been recognised. This has resulted in the development of numerous guidelines on a host of topics which are underpinned by properly appraised evidence for use by clinicians and patients to aid and inform choices in practice [1–2]. One of the key differences between interventions in clinical medicine and public health medicine is that the latter focuses its interventions on populations, of varying sizes, rather than individuals [2]. That being said, the concept of basing recommendations on evidence that is treated in a systematic way using the tools of EBM is logically sound. How best to harness and synthesise evidence from the lower levels of the evidence hierarchy in a way that produces clear advice and recommendations is an important area for discussion currently in the field of public health [3–6].

Furthermore, for new and emerging infectious diseases, and indeed for some established infectious diseases as well, there is not likely to be an extensive body of published literature to draw upon to support specific control measures, e.g. social distancing, contact tracing requirements, etc. Advice will often have to be derived from microbiological and virological principles and field data from surveillance and outbreak investigations. Sometimes advice has to be based on analogy and modelling – but even models to be meaningful have to include a minimum of data, which may be difficult to collect. When empirical data become available to the wider public health community they first appear in expert committee papers and conference presentations, well before peer-reviewed publication, creating a special challenge in identifying that knowledge systematically and quickly.

Definitive scientific evidence is not always available. In a rapidly evolving situation, like the first phase of the H1N1 pandemic, scientific decisions were based on high levels of uncertainty and this lack of certainty clearly frustrated decision-makers at times.

2.3 Methods

The methods employed by the working group consisted of discussion and presentation of different Member States' experiences of providing evidence-based guidance in circumstances when time was short.

An overview of commonly used approaches were presented and discussed along with the elements that were required to produce rapid and robust advice and guidance. This was compared with the methods used to produce standard EB guidance. Tools such as SBAR (Situation, Background, Assessment, Recommendation), which is currently used as part of health protection response to outbreaks of healthcare-associated infections (HAI) in Scotland, and a service currently provided to primary care in Wales to enable rapid access to scientific evidence assessments [7] were tabled and discussed. Modifications to such tools to use them for our purpose were discussed (see Annex 2).

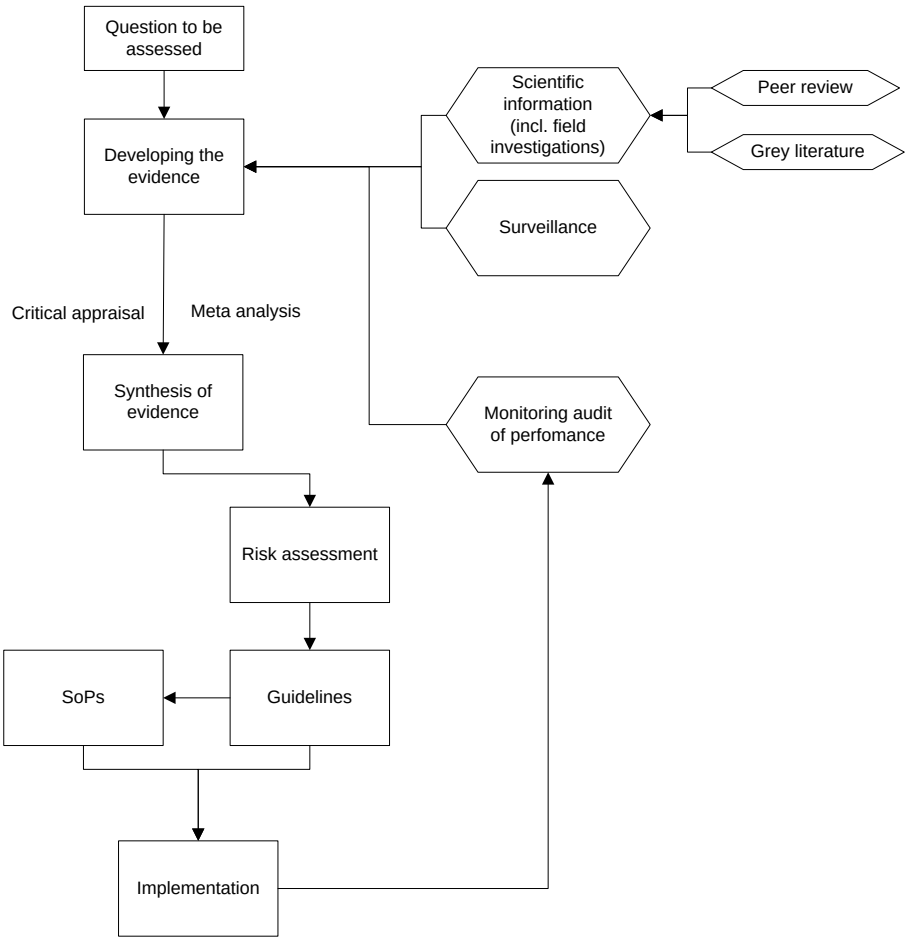
The schema produced and described in this report resulted from thinking through the process of how EB methods may be applied to the public health assessment and input to the management of an outbreak of communicable disease.

2.4 Results and findings

Two schemes are proposed to aid thinking about how evidence-based methods might be applied to acute communicable disease control, the nature of the tasks, and needs to be made explicit and systematic. The first is a standardised causal pathway for infectious diseases that makes explicit the key epidemiological parameters that need to be understood to develop evidence-based advice. The evidence relates to different types of components of this pathway. In terms of the EBM approach, this would closely equate to the approach of NICE to synthesising the different evidence that come from epidemiology and social sciences, which is the use of logic modelling [8–10].

The second schema is an evidence cycle linked to the risk assessment process.

Figure 1 Evidence cycle linked to the risk assessment process



For convenience, the risk assessment process is set out as comprising five stages. The process begins with risk detection and verification, then a rapid risk assessment, appraisal of options for control, development of control measures, policies and advice, followed by implementation, monitoring and evaluation.

Stage 0: Preparatory
Stage 1: Threat detection/verification
Stage 2: Assessment of risk
Stage 3: Developing advice
Stage 4: Implementation
Stage 5: Monitoring and evaluation

We also describe a preparatory stage before stage 1. For each stage we give an explanatory text and propose a list of tools to facilitate the workflow at each stage.

2.4.1 Stage 0: Preparatory

Much can and should be done in periods between communicable disease incidents. Alerting and surveillance systems should be set up and regularly reviewed for fitness for purpose. Criteria and processes for evaluating surveillance systems have been developed [11]. For the priority diseases, summaries of evidence (Cochrane Reviews) should be kept up to date. International cooperation of experts is needed to accomplish this.

Much of the evidence needed for the control of emerging infections must come from field investigations, and outbreak investigations represent a very important source of information. Many recommendations and guidelines for prevention of communicable diseases are based on published outbreak and incident reports. The basic characteristics of specific pathogens (e.g. reproduction number, R0) are frequently derived from observational studies in outbreaks. In addition to careful descriptive epidemiology of cases and incidents, two classic study types are regularly employed for hypothesis testing in outbreak investigations: cohort and case-control studies. The reliability of these investigations is largely dependent upon the strengths and weaknesses in study design, the rigor with which they are conducted, and the appropriateness of the statistical analysis of the data obtained.

Although these observational study designs used in outbreak investigations have been covered by numerous textbooks, it has become obvious that several flaws regularly limit the validity of published outbreak reports. Problems include the identification and selection of potential confounding variables, the definition of cases and controls, or poor and insufficient data analysis. The utility of outbreak investigations and outbreak reports in public health would be greatly improved if a standardised framework for conducting, reporting and auditing outbreak studies was developed and applied [12]. Initiatives like the Consolidated Standards of Reporting Trials (CONSORT) statement have shown that recommendations on the reporting of research can improve reporting quality. Recently, similar initiatives have followed for observational studies, including the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement, the Transparent Reporting of Evaluations of Nonrandomised Designs (TREND) statement, and the Outbreak Reports and Intervention Studies of Nosocomial Infection (ORION) statement. These approaches should be developed also for studies done in epidemic settings.

It seems that there are currently no agreed international standards on how to perform and report outbreak investigations. Elements that should be addressed to allow a transparent review and assessment of new knowledge should include the following:

- detection and confirmation of an outbreak;
- the process of reviewing relevant literature;
- case ascertainment and investigation;
- analysis and reporting of epidemiology, e.g. time, place and persons;
- hypothesis generation and choice of study design;
- the design of a questionnaire;
- the case definition;
- the selection of controls;
- field methods of data collection.

Tools

- Ongoing alert and verification systems (Early Warning and Response System, the EU reporting system for early warnings and response).
- Up-to-date systematic reviews and summaries of evidence (e.g. Cochrane reviews and Attract service for primary care in Wales [7]).
- Quality standards for surveillance and field investigation and reporting (adapt the STROBE and ORION guidelines for reporting outbreaks).

2.4.2 Stage 1: Incident verification

Communicable disease units receive and actively search for rumours and reports of outbreaks and incidents. The critical step at this stage is to recognise the alert signal from the background 'noise'.

The agreed terminology (action in parenthesis) outlining the epidemic intelligence process is:

- signal (filter);
- event (validate);
- validated event (analyse);
- risk (assess).

This process requires rapid communication networks of communicable disease units internationally. Algorithms should be followed to make judgements explicit, with built-in trigger levels that would allow 'stopping rules', so that resources are prioritised efficiently.

Tools

See Executive Summary of rapid risk assessment methodology tool, in Annex 4**.**

2.4.3 Stage 2: Assessment of risk

This stage assumes that there has been an assessment of verified threat and that there is a situation that is known, e.g. SARS or influenza, which requires urgent advice for public health. The assessment needs to consider threats to the population as a whole and to risk groups for more severe disease such as pregnant women, elderly, young and immunecompromised and those at additional risk of exposure, e.g. healthcare workers through occupational exposure.

For rare, new and emerging infections peer-reviewed literature may not have much to say, and greater dependence will be placed on field investigations, ongoing surveillance data and extrapolated evidence and communication with experts in other centres. Rapid access to evidence syntheses is needed, as well as findings from surveillance and field investigations. Currently, no international database of outbreaks is available so that the process of learning from similar incidents in other places is hampered.

Avoiding publication and reporting bias and admitting to gaps in knowledge will be critical to characterise the incident and the risks posed. Therefore, systematic methods for rapid searching and appraisal need to be developed that are appropriate to the time scales involved, which are measured in hours rather than days. This evidence should be organised according to the questions of importance to the public and policymakers.

In order to reduce the risk of bias, reproducible, transparent and explicit risk assessment protocols and algorithms should be followed, and these should explicitly include frameworks for syntheses of different types of evidence in relation to public health questions (e.g. risk of H1N1 to pregnant women at different stages of pregnancy) and admit to gaps and uncertainties in the evidence and alternative explanations of findings. Evidence should be classified (case reports, population surveillance, field investigation) and appraisal of study quality recorded.

Tools

- Protocol for rapid searching and appraisal of peer-reviewed and grey literature needs to be developed or modified from existing search methods to fit purpose.
- Protocol for communication between experts internationally needs to be developed.
- International database on incidents and reports – needs to be agreed and developed.
- Protocol for sharing surveillance data internationally in a rapid fashion.
- Risk assessment framework for synthesis of evidence in relation to public health questions (The version of SBAR is an example, see Annex 3).
- Classification of types of evidence (case reports, population surveillance and field investigation) needs to be developed or modified from existing classifications (assessment of quality of evidence can be done using the GRADE approach. The GRADE approach may be revised through international collaboration to better fit public health questions).

2.4.4 Stage 3: Developing advice

For major incidents, traditionally, governments have convened groups of experts to review and weigh the evidence and to advise ministers on policy and interventions. For example, in the UK, the role of the Spongiform Encephalopathy Advisory Committee was thoroughly reviewed by an independent body (The Phillip Enquiry) and general principles of the conduct for such committees were established. Emphasis was given to openness in dealing with uncertainty.

It is clear from this that even in the acute situation of infectious disease emergencies agreed protocols for developing policy and advice must be followed. Advice will need to explicitly recognise the situational context and

the population groups to which it applies, and address the scope of the advice and time frames considered. Value judgements need to be explicit, taking into account the views of the public. Where evidence of effectiveness of interventions is lacking, it is important to make explicit the principles from which advice is derived. An essential part of developing advice is to clearly state options for interventions and the expected relative merits of different options. Health improvement and decision impact assessment tools might be of relevance. As evidence tables are used to extract and explore the different elements of a scientific study, uncertainty tables can be used to assess the sources of uncertainties at different stages.

Tools

- Guidance on developing advice, including assessment of quality of evidence (explicit reference to advice for other sources).
- Uncertainty tables addressing the stages and steps of uncertainty arising at different levels [13]. There is a need to further develop a practical adapted format of such an uncertainty table for use in a short-term risk assessment for public health.

2.4.5 Stage 4: Implementation

For effective implementation, advice must be framed by requirements of the target groups. This involves consideration of the public perception and communication of risk. Various governments have published guidelines on this [14] . In acute scenarios, the rapidly changing picture and accumulation of intelligence needs to be explained and caveats about interim advice clearly admitted.

Tools

- Checklist of key points to address in risk communication, including risk perception predictor. Media briefing template.

2.4.6 Stage 5: Monitoring and evaluation

Effectiveness of interventions and advice should be monitored. This should include the assessment of degree of acceptability and feasibility of implementation.

Tools

- Rapid audit and lessons learned process.

2.4.7 Next steps

The main focus of the next steps is the completion of the development, modification and adaption of tools and checklists alongside a description of a framework and a methodology to produce rapid guidance in situations of time pressure. This would ideally include the tools described above with worked examples in a ‘work book’ format and which would provide a uniform, consistent methodology for public health and health protection practitioners.

A summary of the tools for the different stages plus examples for modification/adaptation for use in rapid risk assessments is presented in Annex 2.

2.5 References

[1] Lohr KN. Rating the strength of scientific evidence: relevance for quality improvement programs. Int J Qual Health Care. 2004 Feb;16(1):9-18.

[2] Scottish Intercollegiate Guidelines Network. SIGN 50: a guideline developer's handbook. Edinburgh: SIGN; 2008. Available from: http://www.sign.ac.uk/pdf/sign50.pdf.

[3] Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, et al. Grading quality of evidence and strength of recommendations. BMJ [serial on the Internet]. 2004; 328(7454): Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC428525/pdf/bmj32801490.pdf.

[4] Jenicek M, Stachenko S. Evidence-based public health, community medicine, preventive care. Med Sci Monit. 2003 Feb;9(2):SR1-7.

[5] Kunz R, Burnand B, Schunemann HJ. [The GRADE System. An international approach to standardize the graduation of evidence and recommendations in guidelines]. Internist (Berl). 2008 Jun;49(6):673-80.

[6] Tang KC, Choi BC, Beaglehole R. Grading of evidence of the effectiveness of health promotion interventions. J Epidemiol Community Health. 2008 Sep;62(9):832-4.

[7] Public Health Wales. Attract: about. Available from: http://www.attract.wales.nhs.uk/about.aspx.

[8] Rothman KJ, Greenland S. Causation and causal inference in epidemiology. Am J Public Health. 2005;95 Suppl 1:S144-50.

[9] Kelly MP, Stewart E, Morgan A, Killoran A, Fischer A, Threlfall A, et al. A conceptual framework for public health: NICE's emerging approach. Public Health. 2009 Jan;123(1):e14-20.

[10] Kelly M, Morgan A, Ellis S, Younger T, Huntley J, Swann C. Evidence based public health: A review of the experience of the National Institute of Health and Clinical Excellence (NICE) of developing public health guidance in England. Soc Sci Med. 2010 Sep;71(6):1056-62.

[11] German RR, Lee LM, Horan JM, Milstein RL, Pertowski CA, Waller MN. Updated guidelines for evaluating public health surveillance systems: recommendations from the Guidelines Working Group. MMWR Recomm Rep [serial on the Internet]. 2001; 50(RR-13): Available from: http://www.cdc.gov/mmwr/preview/mmwrhtml/rr5013a1.htm.

[12] Dalton CB, Merritt TD, Durrheim DN, Munnoch SA, Kirk MD. A structured framework for improving outbreak investigation audits. BMC Public Health [serial on the Internet]. 2009; 9: Available from: http://www.biomedcentral.com/content/pdf/1471-2458-9-472.pdf.

[13] Guidance for uncertainty assessment and communication: detailed guidance. Utrecht: Utrecht University; 2003. Available from: http://www.nusap.net/downloads/detailedguidance.pdf.

[14] Brownson RC, Fielding JE, Maylahn CM. Evidence-based public health: a fundamental concept for public health practice. Annu Rev Public Health. 2009 Apr 29;30:175-201.

3 The usefulness of EBM grading tools for grading evidence and recommendations for the field of public health / infectious diseases

3.1 Background

There is general agreement that recommendations regarding public health interventions and infectious diseases should be based on the best available evidence and the considered judgement of it, to ensure they do more good than harm. Guidelines based only on a consensus of expert opinion or on unsystematic literature surveys have long been criticised as not reflecting current medical knowledge and being liable to bias [1–2].

Many organisations now report using systematic and transparent methods to develop clinical recommendations, including a growing number of organisations funded by government [3–4]. However, in the field of infectious diseases/public health, it is often the case that guidelines are developed without an explicit evidence-based methodology. Reviews of guidelines produced by different organisations report that they often do not adhere to their own guidelines recommendations [3,5–6]. A study at WHO in 2003–2004 found that systematic reviews and concise summaries of findings were rarely used for developing recommendations. Instead, processes usually relied heavily on experts in a particular specialty, rather than representatives of those who will have to live with the recommendations or on experts in particular methodological areas [7].

A recent analysis of the evidence behind practice guidelines from the Infectious Disease Society of America found that more than half of the recommendations relied solely on expert opinion or anecdotal evidence [8].

Making public health recommendations involve several dimensions:

(i) cognitive (production of knowledge);
(ii) normative (judgement of the value of policies and programmes and their effects);
(iii) instrumental (production of social change).

Corresponding steps in the production of recommendations are:

(i) collection of quantitative and/or qualitative information;
(ii) critical appraisal of this information based on criteria and reference to norms;
(iii) proposals for decision-making.

These steps do not only concern the effectiveness of an intervention but should also consider other domains such as safety, cost and economic evaluation and ethical, organisational, social and legal aspects.

A rapid summary of working group experiences of contact between national public health institutes and organisations performing literature assessment and systematic review indicated that there was generally limited contact, although there are noteworthy exceptions. Some of the organisations in Europe performing literature assessment, systematic reviews and Health Technology Assessment (HTA) reports have limited experience of developing evidence reports or recommendations in public health and infectious diseases. Some of the public health institutes responsible for giving advice regarding infectious diseases do not use explicit methods assessing the evidence and in making the recommendations.

Different methods for assessing the quality of evidence and strength of evidence are available. The US Agency for Healthcare Research and Quality (AHRQ) considered 40 systems until the year 2000 in a systematic review published in 2002 [9]. The Canadian Optimal Prescribing and Utilization Service (COMPUS) expanded and updated the work by AHRQ until the year 2005 and identified 60 evidence grading systems [10]. Weightman et al. [11] carried out a literature review on the methodology for translating findings from public health research evidence into grades of recommendation for interventions in 2005. They selected 37 relevant papers from January 2000–May 2004 retrieved from 16 databases, and included 14 additional papers suggested by experts. The literature review indicated general agreement that the randomised controlled trial (RCT) has the highest internal validity and, where feasible, is the research design of choice when evaluating effectiveness. However, many commentators felt the RCT may be too restrictive for some public health interventions, particularly community-based programmes. In addition, supplementing data from quantitative studies with the results of qualitative research is regarded as key to the successful replication and ultimate effectiveness of interventions. Based on the literature review and consultation with experts, a framework was developed that derives grades of recommendation, incorporating:

- strength of evidence of efficacy based on the research design and the quality and quantity of evidence (the current NICE system); and

- corroborative evidence (from observational and qualitative studies) for the feasibility and likelihood of success of an intervention if implemented in the UK.

This methodology is now being piloted and the results should be considered in our future work.

The variety of methods for assessing the quality of evidence and grading the strength of recommendations is potentially confusing for both developers and users of guidelines and recommendations in the field of public health and infectious diseases. An overview and assessment of the methods available might assist guideline developers in the field of public health to use more rigorous methods in developing recommendations.

The aim of this section of the report is to evaluate the quality and usefulness of existing assessment tools and grading systems for the field of public health/infectious diseases, and to suggest adaptations if necessary.

3.2 Methods

We identified major existing tools and websites and systems for assessing studies and grading the quality of evidence through literature searching, scanning of websites and contact with experts. Articles that summarised the evaluation of existing systems were also included in the search.

The group was interested in examples and methods in current use for grading the quality of evidence, not only relating to intervention effectiveness but also in relation to other issues, for example threat and risk assessment. Implementation and acceptability issues and user perspectives gleaned from qualitative studies.

The challenges of assessing the quality of evidence and making recommendations in the field of infectious diseases were also discussed during project meetings.

3.3 Results

3.3.1 Systems of guideline development and grading of evidence and recommendations

Several organisations have established systems to guide the development of guidelines, 'guidelines for guidelines'. Here we briefly summarise the guidance and experiences of some of the most influential organisations internationally, with a focus on organisations giving recommendations on public health and infectious diseases. We also present the currently most influential systems used.

3.3.2 WHO – 'Improving the use of research evidence in guideline development'

In 2005 the World Health Organization (WHO), the world's leading public health agency, asked its Advisory Committee on Health Research (ACHR) for advice on ways in which WHO could improve the use of research evidence in the development of recommendations, including guidelines and policies. The ACHR established the Subcommittee on the Use of Research Evidence (SURE) to collect background documentation and consult widely with WHO staff, international experts and end users of WHO recommendations to inform its advice to WHO. This resulted in a series of reviews of methods that are used in the development of guidelines, published as a series of articles in Health Research Policy and Systems [12–28]. The articles were collected in the report 'Improving the use of research evidence in guideline development' from the Norwegian Knowledge Centre for the Health Services [29]. This report provides an executive summary of the recommendations to WHO, also relevant for other organisations, regarding guidelines for guidelines.

The authors suggested that 'guidelines for guidelines' should include information and instructions about the following components:

- Priority setting;
- Group composition and consultations;
- Declaration and avoidance of conflicts of interest;
- Group processes;
- Identification of important outcomes;
- Explicit definition of the questions and eligibility criteria;
- Type of study designs for different questions;
- Identification of evidence;
- Synthesis and presentation of evidence;
- Specification and integration of values;
- Making judgments about desirable and undesirable effects;
- Taking account of equity;

- Grading evidence and recommendations;
- Taking account of costs;
- Adaptation, applicability, transferability of guidelines;
- Structure of reports;
- Methods of peer review;
- Planned methods of dissemination & implementation;
- Evaluation of the guidelines [30].

The authors address the question regarding what types of evidence should be used to answer different types of questions [29]:

> *'The most important type of evidence for informing global recommendations is evidence of the effects of the options (interventions or actions) that are considered in a recommendation. This evidence is essential, but not sufficient for making recommendations about what to do. Other types of required evidence are largely context specific.*
> *The study designs to be included in a review should be dictated by the interventions and outcomes being considered. A decision about how broad a range of study designs to consider should be made in relationship to the characteristics of the interventions being considered, what evidence is available, and the time and resources available.*
> *There is uncertainty regarding what study designs to include for some specific types of questions, particularly for questions regarding population interventions, harmful effects and interventions where there is only limited human evidence.*
> *Decisions about the range of study designs to include should be made explicitly.*
> *Great caution should be taken to avoid confusing a lack of evidence with evidence of no effect, and to acknowledge uncertainty.*
> *Expert opinion is not a type of study design and should not be used as evidence. The evidence (experience or observations) that is the basis of expert opinions should be identified and appraised in a systematic and transparent way.'*

When describing causes of disease, spread and impact of disease, other questions are relevant and other lines of evidence should be considered and evaluated. When direct evidence of evaluation of effects of interventions is lacking, we might have to rely on logic modelling with causal pathways based on indirect evidence.

Regarding what criteria should be used to grade evidence and recommendations, the authors recommendations are [29]:

> *'Both the quality of evidence and the strength of recommendations should be graded. The criteria used to grade the strength of recommendations should include the quality of the underlying evidence, but should not be limited to that.*
> *The approach to grading should be one that has wide international support and is suitable for a wide range of different types of recommendations. The Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach, which is currently suggested in the Guidelines for WHO Guidelines, is being used by an increasing number of other organizations internationally. It should be used more consistently by WHO. Further developments of this approach should ensure its wide applicability.*
> *Although there are arguments for and against using the same grading system across a wide range of different types of recommendations, WHO should use a uniform grading system to prevent confusion for developers and users of recommendations.*

WHO Handbook for Guideline Development provides guidance on the development of documents or publications containing WHO recommendations [30]. The term 'guideline' is used to refer to any document containing WHO recommendations.

WHO has published several recommendations both in the field of public health and infectious diseases, and also regarding health systems, using the GRADE approach. The WHO rapid advice guidelines for pharmacological management of sporadic human infection with avian influenza A(H5N1) virus made a strong recommendation to treat H5N1 patients with oseltamivir in part because of the severity of the disease, although the quality of the underlying evidence was rated as very low [31–32]. A few other examples are rapid advice on treatment of tuberculosis in children [33–34] There is an ongoing process between the GRADE working group and WHO on the application of GRADE in vaccine recommendations, including also other public health institutions and ECDC. Considerations about upgrading the value of observational studies and finding a way of expressing the strength of evidence and recommendations that is understandable to the policymakers are important issues to settle.

The working group discussed several issues about the applicability of the GRADE instrument for risk assessments and public health advice. The main concerns and proposed solutions are summarised in a table in Annex 5.

3.3.3 The Grading of Recommendations Assessment, Development and Evaluation Working Group (GRADE)

The Grading of Recommendations Assessment, Development and Evaluation (short GRADE) Working Group began in the year 2000 as an informal collaboration of people with an interest in addressing the shortcomings of present grading systems in healthcare. In a series of international meetings and correspondence over several years, the GRADE Working Group has derived a set of criteria to assess the quality of evidence and the strength of recommendations [35]. Many international organisations have provided input into the development of the approach and more than 50 organisations worldwide endorse or use GRADE [36]. This is the reason why we have focused more on evaluating this specific instrument and it's applicability in a public health setting (see Annex 5).

In the context of making recommendations, the quality of evidence in the GRADE approach reflects the extent to which our confidence in an estimate of the effect is adequate to support a particular recommendation.

Although the quality of evidence might be considered a continuum, the GRADE approach has four levels of the quality of the evidence: high, moderate, low and very low. Systematic reviews of randomised controlled trials start out as high-quality evidence, whereas systematic reviews of observational studies start as low quality. But the quality of evidence is not only determined by the study design. Five factors can lower the quality of evidence, these are:

- serious or very serious limitation to study quality;
- important inconsistency;
- some or major uncertainty about directness;
- imprecise or sparse data; and
- high probability of reporting bias.

Three factors can increase the quality of evidence, these are:

- strong evidence of association;
- evidence of a dose response gradient; and
- all plausible confounders would have reduced the effect.

Systematic reviews of well-designed observational studies may provide moderate- or high-quality evidence if they are upgraded by, for instance, a large effect or evidence of a dose response gradient.

Strength of recommendation according to GRADE is the degree of confidence we can have that adherence to the recommendation will do more good than harm. This is also a continuum, but GRADE has two levels of strength of the recommendations: strong and weak (or conditional).

GRADE separates grading of quality of evidence and strength of recommendations and provides a framework for moving from assessing evidence to making a recommendation. Determinants of the strength of recommendations are: balance between desirable and undesirable effects, quality of the evidence, values and preferences and costs (resource allocation).

The GRADE system has some advantages by including explicit definition and sequential judgments during the grading process; a detailed description of the criteria for the quality of evidence for single outcomes and for the overall quality of the evidence; weighing the relative importance of outcomes; consideration of the balance between health benefits versus harms, burdens and costs; and the development of evidence profiles and summaries of findings. In addition, the GRADE group is supported by an international collaboration [26]. The main limitations and criticism of the GRADE system is connected to its complexity and its applicability in other settings than clinical medicine (see Annex 5 with table of concerns).

Detailed information about the GRADE approach of grading the quality of evidence and the strength of recommendations, with references to relevant publications and presentations, the list of organisations endorsing GRADE and link to a free software to support the use of GRADE, can be found at the website [36] and in the articles describing how to use GRADE [35,37–46].

GRADE has mostly been used in the development of clinical guidelines, but it has also been used in developing recommendations in the field of public health and infectious diseases. There is an ongoing international collaborative effort to apply the GRADE approach to public health and health systems interventions, and it is possible that modifications may be needed to ensure its usefulness for non-clinical interventions [26]. GRADE is a work in progress, continuously being developed, based on feedback and experiences in guideline processes and discussions in the GRADE Working Group.

3.3.4 National Institute for Health and Clinical Excellence (NICE)

The National Institute for Health and Clinical Excellence (NICE) has developed a manual for the development of NICE clinical practice guidelines [36], and one about methods for the development of NICE public health guidance [47]. NICE has also developed a related document on the process of developing public health guidance [48].

For its clinical guidelines, NICE has adopted GRADE methodology in grading the quality of evidence, but it does not grade the strength of the recommendations. NICE has not adopted GRADE for its public health guidance, mainly because of what is considered its limited applicability to the assessment of the broader types of evidence that NICE routinely uses in developing public health guidance (i.e. other than intervention effectiveness studies and including qualitative research about the views and experiences of target populations and practitioners). Individual studies are assessed using standard checklists (one for quantitative intervention studies, one for quantitative correlation studies and one for qualitative studies) to give a rating (++, +, -) that reflects the degree of confidence in the findings of the study. Information is collated in evidence tables and summary 'evidence statements' that reflect the 'strength' of the evidence (number and types of studies, their quality and consistency of findings), the direction and size of effect (where applicable, for intervention and correlation studies) and its applicability. Further details can be found in sections 5.3–5.6 of the manual [47].

While NICE does not grade the strength of recommendations in its public health guidance, these take into account the 'strength' of the evidence and its applicability (see above), as well as the typical effect sizes (where relevant), the importance of the outcomes (including impact on inequalities), trade-offs between harms and benefits, cost effectiveness and other issues (e.g. equality, ethics). Further details can be found in sections 7.2 and 7.3 of the manual [47].

In essence, the NICE public health method for assessing the quality of evidence and process for making judgements about when and how to make recommendations follows the same principles as GRADE, but also allows for a wider set of evidence to be used in decisions about whether recommendations should be made.

3.3.5 Scottish Intercollegiate Guidelines Network (SIGN)

A description of the Scottish Intercollegiate Guidelines Network (SIGN) methodology, together with examples of checklists, evidence tables and considered judgement forms, is given in 'SIGN 50: A guideline developer's handbook' [49].

In 2009, SIGN took the decision to implement the GRADE approach within its guideline development methodology. SIGN has adopted the key principles as set out by the GRADE Working Group, as they stood in June 2010 [50].

SIGN does not offer public health or health protection advice unless a specific question is posed within the context of a clinical guideline, e.g. preventing dental caries in pre-school children. SIGN works closely with the agencies in Scotland who issue such guidance, Heath Protection Scotland and NHS Health Scotland.

As part for the adoption of GRADE principles, SIGN, in collaboration with Andrew Tannahill, Consultant in Public Health Medicine, NHS Health Scotland, has revised the SIGN considered judgment approach for devising recommendations [51]. A key underlying principle of the revised approach is that it preserves the integrity of the evidence and its processing and appraisal in accordance with agreed protocols, without falling into the trap of assuming that strong scientific evidence necessarily implies important recommendations, or that experience-, expertise- and values-based judgements without strong scientific evidence necessarily imply unimportant recommendations. The forms that SIGN has developed for this considered judgment approach for devising recommendations are attached (see Annex 6).

3.3.6 Centers for Disease Control and Prevention (CDC)

The Healthcare Infection Control Practices Advisory Committee (HICPAC) is a US federal advisory committee made up of 14 external infection control and public health experts who provide guidance to the Centers for Disease Control and Prevention (CDC) and the Secretary of the Department of Health and Human Services (DHHS) regarding the practice of healthcare infection prevention and control, strategies for surveillance, and prevention and control of healthcare-associated infections (HAIs) in US healthcare facilities. HICPAC has adopted GRADE in its updated guideline methodology [52].

3.3.7 Other organisations giving guidance on guideline development

In Australia, the National Health and Medical Research Council (NHMRC) has developed standards and procedures for guideline development [53]. Feedback from guideline developers indicated that the previous levels of evidence used by the NHMRC for interventions studies were too restrictive. A 2004 report commissioned by NHMRC identified 18 evidence frameworks that were relevant for evaluation of non-interventional evidence [54]. The Centre for Evidence Based Medicine (CEBM) hierarchy include levels of evidence for assessing questions of therapy/prevention and aetiology/harm, prognosis, diagnosis, differential diagnosis/symptom prevalence, and

economic and decision analysis [55]. NHMRC found the CEBM framework to be the most comprehensive in terms of addressing different types of questions, and used this to model a revised system with additional levels of evidence and grades for recommendations [54,56]. NHMRC has developed and pilot-tested a new framework for guideline development, the FORM approach. A 'body of evidence matrix' contains five factors that impact the strength of a recommendation: the evidence base, consistency, clinical impact, generalisability and applicability in four levels (excellent, good, satisfactory, poor). NHMRC uses four strengths of recommendations, labelled A, B, C and D, linked directly to the 'body of evidence' [57].

US Task Force on Community Preventive Services (USTFCPS) described methods to develop an evidence-based guide to community preventive services in 2000 [58–59]. USTFCPS uses a system in which the quality of the evidence of effectiveness links directly to the strength of the recommendation.

The Infectious Disease Society of America (IDSA) began to employ the use of the GRADE system in new guidelines and guideline updates initiated after October 2008.

The recommendations from the American ACIP (Advisory Committee on Immunization Practices) are based not only on available scientific evidence but also on expertise that comes directly from a diverse group of healthcare providers and public health officials, including professionals from academic medicine, paediatrics, family practice, pharmacy and a member from the non-governmental Immunization Action Coalition. ACIP has adapted the principles of GRADE, but are using narratives to express the quality of evidence instead of high, moderate and low quality evidence [60].

The US Preventive Services Task Force (USPSTF) gives guidance on screening and preventions. The methods and processes for developing guidance are laid out on their website [61]. Included here are revised GRADE definitions (after May 2007) and a chapter on insufficient evidence.

3.4 Articles evaluating existing systems

The GRADE working group started out as an informal collaboration of methodologists and guideline developers, with an interest in addressing the shortcomings of the existing grading systems. Six prominent systems for grading levels of evidence and strength of recommendations were selected and someone familiar with each system prepared a description of each of these [62]. The six grading systems selected, described and assessed were: American College of Chest Physicians (ACCP), Australian National Health and Medical Research Council (NHMRC), Oxford Centre for Evidence-based Medicine (OCEBM), Scottish Intercollegiate Guidelines (SIGN), US Preventive Services Task Force (USPSTF), and US Task Force on Community Preventive Services (USTFCPS). Twelve assessors independently evaluated each system based on 12 criteria to assess the sensibility of the different approaches. Systems used by 51 organisations were compared with these six approaches.

The authors found that there was poor agreement about the sensibility of the six systems. Only the approach of Oxford CEBM was suitable for all four types of questions considered (effectiveness, harm, diagnosis and prognosis). None of the systems was considered usable for all of the target groups considered (professionals, patients and policymakers). The raters found low reproducibility of judgements made using all six systems.

The authors concluded that all of the currently used approaches to grading levels of evidence and the strength of recommendations had important shortcomings [63].

As described above, Merlin et al. also found Oxford CEBM most suitable for non-interventions questions [54].

The Canadian Optimal Medication Prescribing and Utilization Service (COMPUS) identified and evaluated nearly 60 evidence grading systems using the AHRQ grid [9–10]. The highest scoring systems were the GRADE and the SIGN approach [11]. A second round of expert consultation and stakeholder input confirmed the selection of these instruments.

Petitti et al. [64] in an article on insufficient evidence and the update on the methods of the US Preventive Services Task Force (USPSTF), report that after deliberation they decided not to adopt the GRADE approach. Guyatt et al. [65] responded that Petitti and colleagues made three potentially misleading statements about the GRADE approach. Petitti et al., in their response, state that the USPSTF and GRADE use nearly identical criteria to rate the quality of studies that provide information about health benefits and harms. The USPSTF and GRADE both attempt to make their methods and processes transparent and to remain scrupulously free of financial conflicts of interest. The USPSTF wanted further productive dialogue about and convergence of the methods and processes of the USPSTF, GRADE and other authoritative groups.

Bagshaw and Bellomo [66–67] state that the GRADE system represents a considerable improvement from the traditional hierarchies of grading the quality of evidence and strength of recommendations. They suggest, however, to include other aspects such as biological plausibility, reproducibility, generalisability, temporality and coherence.

Kavanagh [68] has criticised the GRADE system, arguing that even though it has evolved through the evidence-based medicine movement, there is no evidence that GRADE itself is reliable. Ansari et al. [69] responded that they see GRADE as a framework uncovering implicit subjectivity and invoking a systematic, explicit, judicious, and transparent approach to interpreting, as opposed to 'capturing' evidence. GRADE reveals how values are assigned to judgments, but GRADE does not dictate what values are assigned simply because it cannot dictate [69].

Ibargoyen-Roteta et al. [70] performed a SWOT (strengths, weaknesses, opportunities and threats) analysis to evaluate an experience of using the GRADE approach to formulate recommendations for a new health technology, comparing GRADE with the SIGN approach that had been used previously. The authors found that application of the GRADE approach allowed recommendations to be formulated and the method to be clarified and made more explicit and transparent. Some challenges were identified, but none of these were specific to GRADE. GRADE was considered to be a more time-consuming method than the SIGN method. An advantage of GRADE was taking into account patient values when defining and grading the relevant outcomes, thereby avoiding any influence from literature precedents. Ibargoyen-Roteta et al. [70] concluded that the GRADE approach could be appropriate for making the recommendation development process for Health Technology Assessment (HTA) reports more explicit.

Barbui et al. [71] describe the challenges in the use and adaptations of the GRADE approach in developing guidelines for the mental health GAP Action Programme in WHO. The authors' experiences suggested that GRADE may be applied as a useful technical framework for synthesising and presenting evidence on the effectiveness of clinical interventions. The authors conclude, however, that the process may be further improved in the following domains: inclusion of non-randomised evidence and evidence that cannot be meta-summarised and analysed; better reproducibility and internal consistency; and consideration of the choice of one among several measures for each outcome to reduce the selection bias [71].

Dahm and Djulbegovic [72] have reviewed the elements of the Australian FORM approach and compared it to other methods. They concluded that the FORM approach offers a methodologically rigorous approach to guidelines development that places particular emphasis on aspects of applicability.

3.5 Guidance for reporting of research studies and checklists for critical assessment

A summary of the best available research evidence is essential, but not sufficient to inform recommendations. To reduce the risk of bias and errors that occur by change, and to facilitate critical appraisal of syntheses of evidence, reviews should be systematic and should explicitly report the methods that were used [19].

Here we describe some tools that are relevant when assessing and developing recommendations in the field of public health and infectious diseases.

3.5.1 Appraisal of Guidelines Research & Evaluation

The AGREE (Appraisal of Guidelines Research & Evaluation) guideline appraisal instrument [73] provides a validated, internationally agreed framework for assessing the quality of clinical practice guidelines. The AGREE instrument considers six different aspects, or domains, of guideline development:

1. Scope and purpose
2. Stakeholder involvement
3. Rigour of development
4. Clarity and presentation
5. Applicability
6. Editorial independence

The domain rigour of development states that, in a good quality guideline, the criteria for selecting the evidence and the methods used for formulating the recommendations should be clearly described and there should be an explicit link between the recommendations and the supporting evidence (read more about the AGREE instrument in Chapter 4)

3.5.2 Critical appraisal of systematic reviews and primary studies of different study designs

A lot of checklists exist that can be used in critical appraisal of different types of studies. The guidance documents of SIGN and NICE are useful resources for such checklists [47,49].

Oxman et al. have assessed the literature on the synthesis and presentation of research evidence [15]. The first of two reviews of different instruments for critically appraising systematic reviews found 20 systems concerned with the appraisal of systematic reviews or meta-analyses, including one scale, 10 checklists and nine guidance

documents [9]. The authors identified seven key domains that they considered important to appraise: study question, search strategy, inclusion and exclusion criteria, data abstraction, study quality, data synthesis and analysis, and funding or sponsorship. The second review used a detailed process to evaluate and select a system and expanded the previous work by AHRQ up until 2005 [10]. They identified approximately 240 quality assessment instruments for systematic reviews, randomised controlled trials and observational studies as well as nearly 50 evidence grading systems. The instruments and systems identified were evaluated by type of study using the AHRQ evaluation grids from the first review, and considering descriptive items for most potential instruments and systems. The highest scoring instruments and systems from each grid represented the proposed selections. The proposed selections were then sent to the same experts that were contacted to review and provide comment during the initial expert consultation. Based on the second expert consultation, the AMSTAR 2005 was selected as the best instrument for appraising systematic reviews [74–75]. A description of the rationale for selecting that instrument is not available.

3.5.3 The EQUATOR Network

The EQUATOR Network is an international initiative that seeks to enhance reliability and value of medical research literature by promoting transparent and accurate reporting of research studies. The EQUATOR website provides links to a comprehensive list of reporting guidelines [76].

The reporting guidelines are mainly aimed at health researchers, but provide also useful information for users of medical literature and can assist critical appraisal of different types of studies.

3.5.4 Guidance for reporting experimental studies

Guidance for reporting randomised controlled trials is provided in the Consolidated Standards of Reporting Trials (CONSORT) statement [77].

Advice on reporting of non-randomised evaluations of behavioural and public health interventions is given in TREND [78]. TREND is a 22-item checklist specifically developed to guide standardised reporting of non-randomised controlled trials. The TREND statement complements the widely adopted CONSORT statement developed for randomised controlled trials. A collective effort in promoting transparent reporting is valuable to improve research synthesis and advance evidence-based recommendations for best practices and policies. All researchers, funding agencies, journal editors and reviewers are encouraged to use the TREND Statement as a guide when designing evaluation studies, reporting evaluation results and reviewing manuscripts for scientific publication, and also assessing these studies when making recommendations.

The TREND statement was first published in a special issue of the American Journal of Public Health in March 2004 [74]. The issue was devoted to evaluation research. This special issue contains a number of papers related to the use of non-randomised or quasi-experimental designs in the evaluation of interventions [79].

The quality of research in hospital epidemiology (infection control) must be improved to be robust enough to influence policy and practice. In order to raise the standards of research and publication, a CONSORT equivalent for these largely quasi-experimental studies has been prepared by the authors of two relevant systematic reviews undertaken for the HTA and the Cochrane Collaboration. The statement was revised following widespread consultation with learned societies, editors of journals and researchers. It consists of a 22-item checklist and a summary table. The emphasis is on transparency to improve the quality of reporting and on the use of appropriate statistical techniques [80].

The statement has been endorsed and welcomed by a number of professional special interest groups and societies, including the Association of Medical Microbiologists (AMM), British Society for Antimicrobial Chemotherapy (BSAC) and the Infection Control Nurses' Association (ICNA) Research and Development Group. Like CONSORT, ORION considers itself a work in progress, which requires ongoing dialogue for successful promotion and dissemination. The statement is, therefore, offered for further public discussion and journals are encouraged to trial it as part of their reviewing and editing process and feedback to the authors.

3.5.5 Guidance for reporting observational studies

STROBE stands for an international, collaborative initiative of epidemiologists, methodologists, statisticians, researchers and journal editors involved in the conduct and dissemination of observational studies, with the common aim of STrengthening the Reporting of OBservational studies in Epidemiology [81–82].

Sorinola et al. [83] reviewed 'instructions to authors' pages of a core collection of 249 journals ('Hague' list). They found that 'instructions to authors' pages provided limited and varied information for preparing a case report, and concluded that there is a need for consensus, and more consistent guidance for authors of case report.

3.5.6 Guidance for reporting diagnostic accuracy studies

Evidence on diagnostic accuracy may also be relevant when developing recommendations for infectious diseases.

STAndards for the Reporting of Diagnostic accuracy studies (STARD) [84–85], Quality Assessment of Diagnostic Accuracy Studies (QUADAS) [86] and Quality appraisal tool for studies of diagnostic reliability (QAREL) [87] are relevant tools regarding assessment of diagnostic studies.

3.5.7 Guidelines for reporting systematic reviews and meta-analyses

Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) provides guidance on how to report and assess systematic reviews [88–89].

Meta-analyses of observational studies (MOOSE) [90,85] proposed a checklist containing specifications for reporting of meta-analyses of observational studies in epidemiology and public health. Use of the checklist should improve the usefulness of meta-analyses for authors, reviewers, editors, readers, and decision-makers.

The Newcastle-Ottawa Scale (NOS) is a scale for assessing the quality of non-randomised studies in meta-analyses. A coding manual is developed to assist the use of NOS [91]. Other relevant resources, including guidance on reporting systematic reviews and meta-analyses, can be found in the Cochrane Handbook for Systematic Reviews of Interventions [92].

3.5.8 Guidance for reporting quality improvement studies

Standards for Quality Improvement Reporting Excellence (SQUIRE) [93] support reporting and assessing reports of quality improvement efforts in healthcare.

3.5.9 Non-randomised studies

We have not been able to identify any study giving reporting guidance or assessing quality of studies on risk assessment relevant for infectious diseases.

The literature on assessing the quality of non-randomised intervention studies is expanding, however [94–96]. It has been shown that it is feasible to develop a checklist that can be used to assess the methodological quality of non-randomised studies, although it was also concluded that '...healthcare policies based upon non-randomised studies or systematic reviews of non-randomised studies may need re-evaluation if the uncertainty in the true evidence base was not fully appreciated when policies were made.' Developing or refining existing quality assessment tools for non-randomised studies is therefore considered to be an important target.

In making recommendations, whether clinical guidelines or guidance for public health and infectious diseases, it is important also to consider implementation issues, such as applicability, factors that can modify the effect of the intervention, economic and socio-cultural factors, and applicability of policies across borders. The quality of this information should also be assessed.

3.5.10 Provision of EB advice in a time-pressured situation

SBAR (Situation, Background, Assessment, Recommendation, see Annex 2) may be a good starting point for the development of a checklist for the provision of EB advice in a time-pressured (and possible evidence-pressure or evolving) situation.

SBAR is an easy to remember mechanism that can be used to frame conversations, especially critical ones, requiring immediate attention and action. It enables clarification on what information should be communicated between members of the team, and how. The tool consists of standardised prompt questions within four sections, to ensure that team members are sharing concise and focused information. The tool helps anticipate the information needed by colleagues and leads assessment.

It was originally used in the military and aviation industries; SBAR was further developed for healthcare and has become commonly used as a tool to aid multidisciplinary work in clinical care to aid communication and clarity and has been promoted as part of patient safety initiatives. SBAR has been used within the field of health protection and specifically in dealing with healthcare-associated infections and outbreaks of infection in hospitals in Scotland [97]. It has become part of an outbreak management process to describe recommendations for action at different stages of the outbreak and to clearly articulate the evidence underpinning the recommendations and subsequently the reason for changing the recommendations as the situation evolved and evidence changed. It helps to provide clarity and transparency to the process of outbreak management.

This method of communication was used to produce quick clear assessments on a number of recommendations for action, which were produced as part of HPS response to the A(H1N1) pandemic in Scotland. The use of the SBAR tool would be aimed particularly at situations when specific questions need to be answered under time pressure

and would enable the process to be detailed with the recommendations. It would act as an audit trail for the process and decisions made.

3.6 Conclusion

Many of the tools required to produce evidence-based recommendations in public health and health protection already exist. Guidance on critically appraising many of the study types that are important to this work is available, although validated tools and checklists to practically undertake this task in the field of public health have not been fully developed.

Much of the evidence that is available to support recommendations in public health and health protection comes from studies that are graded as low or very low quality evidence in the classical EBM hierarchy, and in the GRADE approach.

Some of the challenges are:

- to develop a series of validated checklists for types of evidence where such checklists do not exist, that can be used internationally;
- to come to a consensus on how important evidence from studies that are not included in the classical EB hierarchy is, and where these studies should fit in the hierarchy;
- to develop robust criteria appropriate to public health, health protection and infectious diseases for upgrading and downgrading these studies within the context of GRADE principles, in collaboration with the GRADE Working Group;
- to develop a method of considered judgement that takes into account appropriate values and judgements and contextual issues in public health, health protection and infectious diseases.

3.7 References

[1] Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. Treatments for myocardial infarction. JAMA. 1992;268(2):240-8.

[2] Mulrow CD. Rationale for systematic reviews. BMJ. 1994;309(6954):597-9.

[3] Burgers JS, Grol R, Klazinga NS, Makela M, Zaat J. Towards evidence-based clinical practice: an international survey of 18 clinical guideline programs. Int J Qual Health Care. 2003;15(1):31-45.

[4] Moynihan R, Oxman AD, Lavis JN, Paulsen E. Evidence-informed health policy: using research to make health systems healthier. A review of organizations that support the use of research evidence in developing guidelines, technology assessments, and health policy, for the WHO Advisory Committee on Health Research. Oslo: Knowledge Centre for the Health Services 2008.

[5] Grilli R, Magrini N, Penna A, Mura G, Liberati A. Practice guidelines developed by specialty societies: the need for a critical appraisal. Lancet. 2000;355(9198):103-6.

[6] Shaneyfelt TM, Mayo-Smith MF, Rothwangl J. Are guidelines following guidelines? The methodological quality of clinical practice guidelines in the peer-reviewed medical literature. JAMA. 1999;281(20):1900-5.

[7] Oxman AD, Lavis JN, Fretheim A. Use of evidence in WHO recommendations. Lancet. 2007 Jun 2;369(9576):1883-9.

[8] Lee DH, Vielemeyer O. Analysis of Overall Level of Evidence Behind Infectious Diseases Society of America Practice Guidelines. Archives of Internal Medicine. 2011;171(1):18-22.

[9] West S, King V, Carey TS. Systems to Rate the Strength of Scientific Evidence. Evidence Report/Technology Assessment No. 47 (Prepared by the Research Triangle Institute-University of North Carolina Evidence-based Practice Center under Contract No. 290-97-0011). AHRQ Publication No. 02-E016. Rockville, MD: Agency for Healthcare Research and Quality. April 2002. Rockville: Agency for Healthcare Research and Quality2002.

[10] Canadian Coordinating Office for Health Technology Assessment. Proposed evaluation tools for COMPUS: opportunity of input 2005. Available from: http://www.cadth.ca/media/pdf/COMPUS_Evaluation_Methodology_draft_e.pdf.

[11] Weightman A, Ellis S, Cullum A, Sander L, Turley R. Grading evidence and recommendations for public health interventions: developing and piloting a framework. London: Health Development Agency; 2005. Available from: http://www.nice.org.uk/niceMedia/docs/grading_evidence.pdf.

[12] Boyd EA, Bero LA. Improving the use of research evidence in guideline development: 4. Managing conflicts of interests. Health Res Policy Syst. 2006;4:16.

[13] Edejer TT. Improving the use of research evidence in guideline development: 11. Incorporating considerations of cost-effectiveness, affordability and resource implications. Health Res Policy Syst [serial on the Internet]. 2006; 4: Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1764011/pdf/1478-4505-4-23.pdf.

[14] Fretheim A, Schunemann HJ, Oxman AD. Improving the use of research evidence in guideline development: 5. Group processes. Health Res Policy Syst [serial on the Internet]. 2006; 4: Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1702534/pdf/1478-4505-4-17.pdf.

[15] Fretheim A, Schunemann HJ, Oxman AD. Improving the use of research evidence in guideline development: 15. Disseminating and implementing guidelines. Health Res Policy Syst [serial on the Internet]. 2006; 4: Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1712226/pdf/1478-4505-4-27.pdf.

[16] Fretheim A, Schunemann HJ, Oxman AD. Improving the use of research evidence in guideline development: 3. Group composition and consultation process. Health Res Policy Syst [serial on the Internet]. 2006; 4: Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1702349/pdf/1478-4505-4-15.pdf.

[17] Oxman AD, Fretheim A, Schunemann HJ. Improving the use of research evidence in guideline development: introduction. Health Res Policy Syst [serial on the Internet]. 2006; 4: Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1684247/pdf/1478-4505-4-12.pdf.

[18] Oxman AD, Schunemann HJ, Fretheim A. Improving the use of research evidence in guideline development: 12. Incorporating considerations of equity. Health Res Policy Syst [serial on the Internet]. 2006; 4: Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1702352/pdf/1478-4505-4-24.pdf.

[19] Oxman AD, Schunemann HJ, Fretheim A. Improving the use of research evidence in guideline development: 8. Synthesis and presentation of evidence. Health Res Policy Syst [serial on the Internet]. 2006; 4: Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1702353/pdf/1478-4505-4-20.pdf.

[20] Oxman AD, Schunemann HJ, Fretheim A. Improving the use of research evidence in guideline development: 7. Deciding what evidence to include. Health Res Policy Syst [serial on the Internet]. 2006; 4: Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1702350/pdf/1478-4505-4-19.pdf.

[21] Oxman AD, Schunemann HJ, Fretheim A. Improving the use of research evidence in guideline development: 16. Evaluation. Health Res Policy Syst [serial on the Internet]. 2006; 4: Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1702533/pdf/1478-4505-4-28.pdf.

[22] Oxman AD, Schunemann HJ, Fretheim A. Improving the use of research evidence in guideline development: 2. Priority setting. Health Res Policy Syst [serial on the Internet]. 2006; 4: Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1702532/pdf/1478-4505-4-14.pdf.

[23] Oxman AD, Schunemann HJ, Fretheim A. Improving the use of research evidence in guideline development: 14. Reporting guidelines. Health ResPolicy Syst [serial on the Internet]. 2006; 4: Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1702351/pdf/1478-4505-4-26.pdf.

[24] Schunemann HJ, Fretheim A, Oxman AD. Improving the use of research evidence in guideline development: 10. Integrating values and consumer involvement. Health ResPolicy Syst [serial on the Internet]. 2006; 4: Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1697808/pdf/1478-4505-4-22.pdf.

[25] Schunemann HJ, Fretheim A, Oxman AD. Improving the use of research evidence in guideline development: 13. Applicability, transferability and adaptation. Health ResPolicy Syst [serial on the Internet]. 2006; 4: Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1712227/pdf/1478-4505-4-25.pdf.

[26] Schunemann HJ, Fretheim A, Oxman AD. Improving the use of research evidence in guideline development: 9. Grading evidence and recommendations. Health ResPolicy Syst [serial on the Internet]. 2006; 4: Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1697807/pdf/1478-4505-4-21.pdf.

[27] Schunemann HJ, Fretheim A, Oxman AD. Improving the use of research evidence in guideline development: 1. Guidelines for guidelines. Health ResPolicy Syst [serial on the Internet]. 2006; 4: Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1665445/pdf/1478-4505-4-13.pdf.

[28] Schunemann HJ, Oxman AD, Fretheim A. Improving the use of research evidence in guideline development: 6. Determining which outcomes are important. Health ResPolicy Syst [serial on the Internet]. 2006; 4: Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1687184/pdf/1478-4505-4-18.pdf.

[29] Oxman A, Schünemann HJ, Fretheim A, SURE. Improving the use of research evidence in guideline development. Oslo: The Norwegian Knowledge Centre for the Health Services; Report no 11-2007. Available from: http://www.kunnskapssenteret.no/Publikasjoner/894.cms

[30] World Health Organization. WHO handbook for guideline development. Geneva: WHO; 2010. Available from: http://www.who.int/hiv/topics/mtct/grc_handbook_mar2010_1.pdf.

[31] Schunemann HJ, Hill SR, Kakad M, Vist GE, Bellamy R, Stockman L, et al. Transparent development of the WHO rapid advice guidelines. PLoS Med [serial on the Internet]. 2007; 4(5): Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1877972/pdf/pmed.0040119.pdf.

[32] Schunemann HJ, Hill SR, Kakad M, Bellamy R, Uyeki TM, Hayden FG, et al. WHO Rapid Advice Guidelines for pharmacological management of sporadic human infection with avian influenza A(H5N1) virus. Lancet InfectDis. 2007;7(1):21-31.

[33] World Health Organization. Rapid advice: treatment of tuberculosis in children. Geneva: WHO; 2010. Available from: http://whqlibdoc.who.int/publications/2010/9789241500449_eng.pdf.

[34] World Health Organization. Treatment of tuberculosis guidelines. Geneva: WHO; 2010. Available from: http://whqlibdoc.who.int/publications/2010/9789241547833_eng.pdf.

[35] Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ. 2008 Apr 26;336(7650):924-6.

[36] GRADE Working Group. GRADE: welcome. Available from: http://www.gradeworkinggroup.org/.

[37] Guyatt GH, Oxman AD, Kunz R, Vist GE, Falck-Ytter Y, Schunemann HJ. What is “quality of evidence” and why is it important to clinicians? BMJ. 2008 May 3;336(7651):995-8.

[38] Guyatt GH, Oxman AD, Kunz R, Falck-Ytter Y, Vist GE, Liberati A, et al. Going from evidence to recommendations. BMJ. 2008 May 10;336(7652):1049-51.

[39] Guyatt GH, Oxman AD, Kunz R, Jaeschke R, Helfand M, Liberati A, et al. Incorporating considerations of resources use into grading recommendations. BMJ. 2008 May 24;336(7654):1170-3.

[40] Schunemann HJ, Oxman AD, Brozek J, Glasziou P, Jaeschke R, Vist GE, et al. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. BMJ. 2008 May 17;336(7653):1106-10.

[41] Jaeschke R, Guyatt GH, Dellinger P, Schunemann H, Levy MM, Kunz R, et al. Use of GRADE grid to reach decisions on clinical practice guidelines when consensus is elusive. BMJ. 2008;337:a744.

[42] Guyatt GH, Oxman AD, Schunemann HJ, Tugwell P, Knottnerus A. GRADE guidelines: a new series of articles in the Journal of Clinical Epidemiology. J Clin Epidemiol. 2011 Apr;64(4):380-2.

[43] Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol. 2011 Apr;64(4):383-94.

[44] Guyatt GH, Oxman AD, Kunz R, Atkins D, Brozek J, Vist G, et al. GRADE guidelines: 2. Framing the question and deciding on important outcomes. J Clin Epidemiol. 2011 Apr;64(4):395-400.

[45] Balshem H, Helfand M, Schunemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol. 2011 Apr;64(4):401-6.

[46] Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, et al. GRADE guidelines: 4. Rating the quality of evidence – study limitations (risk of bias). J Clin Epidemiol. 2011 Apr;64(4):407-15.

[47] National Institute for Health Clinical Excellence. Methods for the development of NICE public health guidance (second edition). London: NICE; 2009. Available from: http://www.nice.org.uk/media/2FB/53/PHMethodsManual110509.pdf.

[48] National Institute for Health and Clinical Excellence. The NICE public health guidance development process: an overview for stakeholders, including public health practitioners, policy makers and the public (second edition). London: NICE; 2009. Available from: http://www.nice.org.uk/media/F19/70/PHProcessGuide2009.pdf.

[49] Scottish Intercollegiate Guidelines Network. SIGN 50: a guideline developer's handbook. Edinburgh: SIGN; 2008. Available from: http://www.sign.ac.uk/pdf/sign50.pdf.

[50] Scottish Intercollegiate Guidelines Network. Applying the GRADE methodology to SIGN guidelines: core principles 2010. Available from: http://www.sign.ac.uk/pdf/gradeprincipals.pdf.

[51] Tannahill A. Beyond evidence – to ethics: a decision-making framework for health promotion, public health and health improvement. Health Promot Int. 2008 Dec;23(4):380-90.

[52] Umscheid CA, Agarwal RK, Brennan PJ. Updating the guideline methodology of the Healthcare Infection Control Practices Advisory Committee (HICPAC). [Atlanta]: Centers for Disease Control and Prevention. Department of Health & Human Services; 2010. Available from: http://www.cdc.gov/hicpac/pdf/guidelines/2009-10-29HICPAC_GuidelineMethodsFINAL.pdf.

[53] The National Health and Medical Research Council. NHMRC standards and procedures for externally developed guidelines. Canberra: NHMRC; 2007. Available from: http://www.nhmrc.gov.au/_files_nhmrc/file/publications/synopses/nh56.pdf.

[54] Merlin T, Weston A, Tooher R. Extending an evidence hierarchy to include topics other than treatment: revising the Australian 'levels of evidence'. BMC Med Res Methodol [serial on the Internet]. 2009; 9: Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2700132/pdf/1471-2288-9-34.pdf.

[55] Phillips B, Ball C, Sackett D, Badenoch D, Straus S, Haynes B, et al. Oxford Centre for Evidence-Based Medicine – levels of evidence [Website] Oxford: Centre for Evidence-based Medicine; 2009 [updated 01 February 2011]; Available from: http://www.cebm.net/index.aspx?o=1025.

[56] The National Health and Medical Research Council. NHMRC additional levels of evidence and grades for recommendations for developers of guidelines – Stage 2 Pilot Canberra: NHMRC; 2007; Available from: http://www.nhmrc.gov.au/guidelines/consult/consultations/add_levels_grades_dev_guidelines2.htm.

[57] Hillier S, Grimmer-Somers K, Merlin T, Middleton P, Salisbury J, Tooher R, et al. FORM: An Australian method for formulating and grading recommendations in evidence-based clinical guidelines. BMC Med Res Methodol [serial on the Internet]. 2011; 11(23): Available from: http://www.biomedcentral.com/content/pdf/1471-2288-11-23.pdf.

[58] Briss PA, Brownson RC, Fielding JE, Zaza S. Developing and using the Guide to Community Preventive Services: lessons learned about evidence-based public health. Annu Rev Public Health. 2004;25:281-302.

[59] Briss PA, Zaza S, Pappaioanou M, Fielding J, Wright-De Aguero L, Truman BI, et al. Developing an evidence-based Guide to Community Preventive Services – methods. The Task Force on Community Preventive Services. Am J Prev Med. 2000 Jan;18(1 Suppl):35-43.

[60] Advisory Committee on Immunization Practices. General recommendations on immunization. MMWR [serial on the Internet]. 2011; 60(2): Available from: http://www.cdc.gov/mmwr/pdf/rr/rr6002.pdf.

[61] U.S. Preventive Services Task Force. Methods and Processes. [updated Current as of January 2011]; Available from: http://www.uspreventiveservicestaskforce.org/methods.htm.

[62] Atkins D, Eccles M, Flottorp S, Guyatt GH, Henry D, Hill S, et al. Systems for grading the quality of evidence and the strength of recommendations I: critical appraisal of existing approaches The GRADE Working Group. BMC Health Serv Res [serial on the Internet]. 2004; 4(1): Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC545647/pdf/1472-6963-4-38.pdf.

[63] Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, et al. Grading quality of evidence and strength of recommendations. BMJ [serial on the Internet]. 2004; 328(7454): Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC428525/pdf/bmj32801490.pdf.

[64] Petitti DB, Teutsch SM, Barton MB, Sawaya GF, Ockene JK, DeWitt T. Update on the methods of the U.S. Preventive Services Task Force: insufficient evidence. Ann Intern Med. 2009;150(3):199-205.

[65] Guyatt GH, Helfand M, Kunz R. Comparing the USPSTF and GRADE approaches to recommendations. Ann Intern Med. 2009 Sep 1;151(5):363; author reply -4.

[66] Bagshaw SM, Bellomo R. The need to reform our assessment of evidence from clinical trials: A commentary. Philosophy, Ethics, and Humanities in Medicine. 2008;3(1).

[67] Bellomo R, Bagshaw SM. Evidence-based medicine: classifying the evidence from clinical trials – the need to consider other dimensions. Crit Care [serial on the Internet]. 2006; 10(5): Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1751050/pdf/cc5045.pdf.

[68] Kavanagh BP. The GRADE system for rating clinical guidelines. PLoS Med [serial on the Internet]. 2009; 6(9): Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2735782/pdf/pmed.1000094.pdf.

[69] Ansari MT, Tsertsvadze A, Moher D. Grading quality of evidence and strength of recommendations: a perspective. PLoS Med [serial on the Internet]. 2009; 6(9): Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2735783/pdf/pmed.1000151.pdf.

[70] Ibargoyen-Roteta N, Gutiérrez-Ibarluzea I, Rico-Iturrioz R, López-Argumedo M, Reviriego-Rodrigo E, Cabriada-Nuño JL, et al. The GRADE approach for assessing new technologies as applied to apheresis devices in ulcerative colitis. Implementation Science. 2010;5(1).

[71] Barbui C, Dua T, van Ommeren M, Yasamy MT, Fleischmann A, Clark N, et al. Challenges in developing evidence-based recommendations using the grade approach: The case of mental, neurological, and substance use disorders. PLoS Medicine [serial on the Internet]. 2010; 7(8): Available from: http://www.plosmedicine.org/article/info%3Adoi%2F10.1371%2Fjournal.pmed.1000322.

[72] Dahm P, Djulbegovic B. The Australian 'FORM' approach to guideline development: The quest for the perfect system. BMC Med Res Methodol [serial on the Internet]. 2011; 11: Available from: http://www.biomedcentral.com/content/pdf/1471-2288-11-17.pdf.

[73] Appraisal of Guidelines Research & Evaluation. Welcome to the AGREE Enterprise website. Available from: http://www.agreetrust.org/.

[74] Shea BJ, Hamel C, Wells GA, Bouter LM, Kristjansson E, Grimshaw J, et al. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. Journal of Clinical Epidemiology. 2009;62(10):1013-20.

[75] Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol [serial on the Internet]. 2007; 7: Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1810543/pdf/1471-2288-7-10.pdf.

[76] EQUATOR Network. About EQUATOR. Available from: http://www.equator-network.org/about-equator/.

[77] Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMC Med [serial on the Internet]. 2010; 8: Available from: http://www.biomedcentral.com/content/pdf/1741-7015-8-18.pdf.

[78] Des Jarlais DC, Lyles C, Crepaz N. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. Am J Public Health [serial on the Internet]. 2004; 94(3): Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1448256/pdf/0940361.pdf.

[79] American Journal of Public Health, 2004; 94(3): Available from: http://ajph.aphapublications.org/content/vol94/issue3/index.dtl.

[80] Stone SP, Cooper BS, Kibbler CC, Cookson BD, Roberts JA, Medley GF, et al. The ORION statement: guidelines for transparent reporting of outbreak reports and intervention studies of nosocomial infection. Lancet Infect Dis. 2007 Apr;7(4):282-8.

[81] Vandenbroucke JP, von Elm E, Altman DG, Gotzsche PC, Mulrow CD, Pocock SJ, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. PLoS Med [serial on the Internet]. 2007 2020496]; 4(10): Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2020496/pdf/pmed.0040297.pdf.

[82] von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. PLoS Med [serial on the Internet]. 2007; 4(10): Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2020495/pdf/pmed.0040296.pdf.

[83] Sorinola O, Olufowobi O, Coomarasamy A, Khan KS. Instructions to authors for case reporting are limited: a review of a core journal list. BMC Med Educ [serial on the Internet]. 2004; 4: Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC400744/pdf/1472-6920-4-4.pdf.

[84] Bossuyt PM, Reitsma JB. The STARD initiative. Lancet. 2003 Jan 4;361(9351):71.

[85] Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. BMJ [serial on the Internet]. 2003; 326(7379): Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1124931/pdf/41.pdf.

[86] Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol [serial on the Internet]. 2003; 3: Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC305345/pdf/1471-2288-3-25.pdf.

[87] Lucas NP, Macaskill P, Irwig L, Bogduk N. The development of a quality appraisal tool for studies of diagnostic reliability (QAREL). J Clin Epidemiol. 2010 Aug;63(8):854-61.

[88] Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med [serial on the Internet]. 2009; 6(7): Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2707010/pdf/pmed.1000100.pdf.

[89] Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med [serial on the Internet]. 2009; 6(7): Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2707599/pdf/pmed.1000097.pdf.

[90] Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA. 2000 Apr 19;283(15):2008-12.

[91] Ottawa Hospital Research Institute. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. Available from: http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp.

[92] The Cochrane Collaboration. Cochrane handbook for systematic reviews of interventions. Version 5.1.0. [updated March 2011] 2011. Available from: http://www.cochrane-handbook.org/.

[93] Davidoff F, Batalden P, Stevens D, Ogrinc G, Mooney S. Publication guidelines for quality improvement in health care: evolution of the SQUIRE project. Qual Saf Health Care. 2008 Oct;17 Suppl 1:i3-9.

[94] Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. J Epidemiol Community Health. 1998 Jun;52(6):377-84.

[95] Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F, et al. Evaluating non-randomised intervention studies. Health Technol Assess. 2003;7(27):iii-x, 1-173.

[96] Pocock SJ, Collier TJ, Dandreo KJ, de Stavola BL, Goldman MB, Kalish LA, et al. Issues in the reporting of epidemiological studies: a survey of recent practice. BMJ. 2004 Oct 16;329(7471):883.

[97] Institute for Innovation and Improvement. SBAR - situation-background-assessment-recommendation. Available from: http://www.institute.nhs.uk/quality_and_service_improvement_tools/quality_and_service_improvement_tools/sbar_-_situation_-_background_-_assessment_-_recommendation.html.

4 Assessing and assuring quality to guideline development for health protection and control of communicable diseases

4.1 Introduction

> *'A guideline which fulfils all the institute's requirements is like the Holy Grail: worth striving for, but unattainable by mere mortals.'*
> (Gene Feder, St Bartholomew's and the Royal London Medical College, 1993)

Evidence-based guidelines are considered essential instruments to improve the quality, appropriateness, and cost-effectiveness of healthcare [1]. Following the immense popularity that clinical guidelines gained in the 90s, leading communicable disease control organisations have also embarked, in the last decade, in the systematic development of evidence-based guidelines.

As tools for improvement, evidence-based guidelines need to be understood and supported in their specific area of practice. The evidence available in each specific context, the research that simultaneously expands both the knowledge and the uncertainty, and the experiential evidence that practitioners bring, have to blend in a particular way, so recommendations, policies and, ultimately their implementation, work effectively.

> *'Scientific evidence and clinical judgment can be systematically combined to produce clinically valid, operational recommendations for appropriate care that can and will be used to persuade clinicians, patients, and others to change their practices in ways that lead to better health outcomes (...)' [2].*

Guidelines are, however, not absolute strategies for improving practice, and enthusiasm for them must be tempered with caution. For practitioners to adopt evidence-based guidelines, they must recognise, understand and agree with the advice given, feel confident about their abilities, and overcome the inertia of previous habits [3]. One important reason that prevents evidence-based guidelines from taking root in practice is that, despite the efforts put into their development, guidelines may provide ambiguous or inconsistent recommendations and have errors or omissions [4].

As guideline quality plays an important role in their credibility, professional bodies and institutions with clearinghouse functions constantly work to improve their guideline development programmes and the quality of their products. Nonetheless, there is still considerable room for improvement, as weak methods or the lack of quality control over methods may still result in low-quality guidelines that could potentially cause confusion or misdirection [5,6].

Approval by those who provide care, by those responsible for monitoring care in the public interest, and by those who understand cost-effective implications, are means of quality assuring the validity of evidence-based guidelines. Other more formal mechanism of validation, however, would perhaps follow agreed measures of the methodological quality of guidelines [7].

This section of the report aims to identify predictors of high quality for evidence-based guidelines for public health services with competences in communicable disease control. These will be important to guideline developers, to help them produce high-quality guidelines, as well as to guidelines users, to help them identify credible guidelines.

4.2 Aims and objectives

The ECDC working group was set up to look at guidelines quality agreed that the main aim of their work would be identifying a set of quality criteria that would be necessary to validate guidelines for communicable disease prevention and control. This aims to help qualifying guidelines development methods as well as guideline documents, and by doing this, this paper would help advise and support guideline development groups (GDG) or clearinghouse bodies accountable to international, national or local institutions with communicable disease control competencies.

The objectives of this working group would therefore address how to:

1. Identify, adjust and/or develop the necessary set of criteria to validate Guidelines for public health. This objective may imply that, depending on the circumstances in practice (e.g. time available to formulate recommendations), the necessary and/or the absolute levels of methods quality may also differ.

2. Validate a consistent framework to Guideline Development that would fit well into any circumstance in which public health advice is needed, with special attention to situations where time to respond is limited.

4.3 Background

4.3.1 Evidence-based practice and the role of guidelines

Evidence-based medicine has been defined as '…the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients' [8].

As the practice of EBM is continuing to evolve in all areas of medicine, public health professionals face a growing pressure to support their decisions with solid evidence.

Clinical practice guidelines are '…systematically developed statements to assist practitioner and patient decisions about appropriate healthcare for specific clinical circumstances' [9]. They aim to improve the effectiveness, quality, efficiency and safety of care, by combining the best available evidence from the latest research with clinical expertise and patient preferences [10]. Evidence-based public health could be defined as integrating the best available evidence with the knowledge and considered judgements from stakeholders and experts to improve health and protect the population from infectious and environmental hazards.

This implies explicitly and systematically displaying the types and strengths and weaknesses of the evidence and the judgments behind a piece of advice or a decision made. Public health guidelines are thus systematically developed statements to assist public health practitioners and policymakers to take decisions about appropriate healthcare, prevention and control of diseases on a population level.

4.3.2 Characteristics of effective guidelines

Guidelines are effective if they lead to changes in practice and improvements in patient outcomes [11]. Numerous studies have investigated the factors that influence the effectiveness of clinical practice guidelines [10,12–19]. These suggest that the characteristics of effective guidelines include:

- the quality of the development methods;
- the transparency of the methods and processes;
- the use and strength of supporting evidence;
- the presentation and format of the content; and
- the inclusion of specific recommendations.

4.3.3 Guideline appraisal instruments

There have been several recent reviews of guideline appraisal instruments. Graham et al. [20], identified 13 appraisal instruments and thematically grouped their quality criteria into 44 'items' across 10 'attributes', i.e. validity; reliability/reproducibility; clinical applicability; clinical flexibility; multidisciplinary process; clarity; scheduled review; dissemination; implementation; and evaluation. A content analysis revealed that only one of these – the Cluzeau instrument [21] – included at least one item for each of the 10 attributes, and overall it addressed 28 items. Furthermore, it was one of only two instruments that had been validated via inter-rater reliability testing. However, Graham et al. [20] concluded that there was insufficient evidence to support the exclusive use of any single instrument.

Vlayen et al. [22] updated the Graham et al. [20] review to include appraisal instruments published up to October 2003. They found a total of 24 instruments and categorised these according to their country of origin, number of quality criteria, scoring system, and the extent to which they had been validated. Based on the work of the Institute of Medicine [2], Vlayen et al. [22] thematically grouped the criteria from the instruments into 50 items across the same 10 guideline attributes used by Graham et al. [20]. In addition to the Cluzeau instrument, they found two further instruments that addressed each of the guideline attributes [23,24]; however, neither of them had been validated.

Vlayen et al. [22] identified two other instruments based on the Cluzeau instrument. One of these – the AGREE instrument [7] – uses a numerical scale and has also been validated (see Section 4.3.4). Vlayen et al. [22] concluded that the AGREE instrument has the most potential to serve as a basis of an appraisal instrument for clinical pathways (the reason for their review).

Finally, Oxman et al. [25] reviewed the literature on evaluating guidelines and recommendations, including their quality, whether they are likely to be up-to-date, and their implementation. This was undertaken as part of a series of reviews for the WHO Advisory Committee on Health Research. Oxman et al. [25] concluded that WHO should use the AGREE instrument or a similar checklist to ensure that guidelines are routinely reviewed. They recommended that the checklist should be adapted and tested for use with a range of WHO recommendations, including public health and health policy. It should also include questions about equity and other issues of particular importance for WHO guidelines. In addition, Oxman et al. [25] recommended that WHO guidelines should be regularly reviewed to determine if they need to be updated.

4.3.4 The AGREE instrument – guideline development and appraisal

The AGREE (Appraisal of Guidelines for Research and Evaluation) instrument was developed to appraise the process of guideline development and how well this is reported. It was developed by an international group of researchers and guideline developers (The AGREE Collaboration) to assess guideline quality, defined as '...confidence that potential biases of guideline development have been addressed adequately and that the recommendations are both internally and externally valid, and are feasible for practice' [7].

The original AGREE instrument was updated in 2009 and includes a new user's manual with detailed assessment criteria and guidance on where to find information and how to rate [26,27]. The AGREE II instrument is a refinement of the original AGREE instrument, and among other changes it is less clinical and better suited for public health guidelines. The purpose of the AGREE II instrument is to:

- assess the quality of guidelines;
- provide a methodological strategy for the development of guidelines; and
- inform what information and how information ought to be reported in guidelines.

The AGREE II is concerned with different aspects of the guideline development process and its reporting across six domains:

1) **Scope and purpose**: the overall aim of the guideline, the specific health questions, and the target population.
2) **Stakeholder involvement**: the extent to which the guideline was developed by the appropriate stakeholders and represents the views of its intended users.
3) **Rigor of development**: the process used to gather and synthesise the evidence, the methods to formulate the recommendations, and to update them.
4) **Clarity and presentation**: the language, structure, and format of the guideline.
5) **Applicability**: the likely barriers and facilitators to implementation, strategies to improve uptake, and resource implications of applying the guideline.
6) **Editorial independence**: the formulation of recommendations not being unduly biased with competing interests.

Each domain is scored using several items (criteria). There are a total of 23 items to be scored on a 7-point Likert scale by at least two (preferably four) independent observers (see Table 1). The appraisal may take each user an average of 90 minutes to complete; therefore, it is designed as a thorough and robust assessment rather than a 'quick check' of the quality of the guideline being assessed. The domain scores are calculated by summing all of the item scores within a domain and standardising the total as a percentage of the maximum possible score for that domain. The AGREE II instrument does not set a minimum score as an indication of quality for each domain and the domain scores are not intended to be added together to give a total score. However, the user is encouraged to rate the overall quality of the guideline and whether or not they would recommend it for use (with or without adaptations).

Table 1 The AGREE II instrument

Domain	Item
Scope and purpose	1. The overall objective(s) of the guideline is (are) specifically described.
	2. The health question(s) covered by the guideline is (are) specifically described.
	3. The population (patients, public, etc.) to whom the guideline is meant to apply is specifically described.
Stakeholder involvement	4. The guideline development group includes individuals from all the relevant professional groups.
	5. The views and preferences of the target population (patients, public, etc.) have been sought.
	6. The guideline has been piloted among end users.
Rigour of development	7. Systematic methods were used to search for evidence.
	8. The criteria for selecting the evidence are clearly described.
	9. The strengths and limitations of the body of evidence are described.
	10. The methods for formulating the recommendations are clearly described.
	11. The health benefits, side effects and risks have been considered in formulating the recommendations.
	12. There is an explicit link between the recommendations and the supporting evidence.
	13. The guideline has been externally reviewed by experts prior to its publication.
	14. A procedure for updating the guideline is provided.
Clarity of presentation	15. The recommendations are specific and unambiguous.
	16. The different options for management of the condition or health issue are clearly presented.
	17. Key recommendations are easily identifiable.
Applicability	18. The guideline describes facilitators and barriers to its application.
	19. The guidelines provides advice and/or tools on how the recommendations can be put into practice.
	20. The potential resource implications of applying the recommendations have been considered.
	21. The guideline presents monitoring and/ or auditing criteria.
Editorial indepen-dence	22. The views of the funding body have not influenced the content of the guideline.
	23. Competing interests of guideline development members have been recorded and addressed.

4.3.5 Limitations of the AGREE II instrument

As discussed above, the AGREE instrument is intended to assess the quality of a guideline's development methods and its reporting, not the quality of its actual content.

Furthermore, the AGREE II instrument states that a recommendation should provide '...concrete and precise description of which option is appropriate in which situation and in what population group, as informed by the body of evidence' [28]. However, although the instrument includes an item about the overall specificity of a guideline's recommendations, it is not possible to appraise each recommendation within a guideline.

Therefore, a limitation of the AGREE II instrument – common to all guideline appraisal instruments – is that it does not evaluate the quality of the content and, more specifically, the evidence supporting the recommendations [22]. To perform such a task requires a more in-depth analysis of the cited studies and an examination of the methods used to appraise and synthesise the evidence to formulate recommendations.

4.3.6 Application of the AGREE instrument

The original AGREE instrument was validated and applied to guidelines in a wide range of clinical practice areas. It has been adopted into the guidelines programmes of several national agencies, including NICE in England, the Scottish Intercollegiate Guidelines Network (SIGN), AZQ in Germany and the New Zealand Guidelines Group (NZGG). The World Health Organization (WHO) uses the instrument to assess its own guidelines. This testifies to the fact that it is internationally regarded as the 'gold standard' for clinical practice guidelines.

However, there are relatively few examples of the AGREE instrument being used to appraise or guide the development of public health guidance and, more specifically, communicable disease guidelines. The Centre for Public Health Excellence at NICE, in England, has based its methods and processes on the AGREE criteria in order to produce over 30 pieces of public health guidance since 2005 [29,30]. However, this has not been without its methodological challenges [31]. In any case, few of these guidelines are directly relevant to communicable diseases and so it is not possible to say with any certainty whether the AGREE criteria can be applied.

In relation to the applicability of the AGREE II instrument to communicable disease guidelines, there are likely additional considerations that need to be taken into account. For example, given that these guidelines may on occasion impact on civil liberties, it is imperative that they are developed with the utmost transparency and democratic accountability.

4.4 Method

In order to identify the necessary set of criteria relevant to guidelines for public health and health protection, a pilot study was conducted to assess a sample of two guidelines focused on areas of importance for communicable disease control. The documents chosen for this study were:

- HPN/HPS guideline for the control of measles incidents and outbreaks in Scotland [32]; and
- HPA management and control of group A streptococcal infections [33].

4.4.1 Assessment of the AGREE II instrument for communicable disease guidelines

The AGREE II instrument was used to appraise these guidelines. Six independent appraisers assessed each guideline. Members of the guideline development group were not eligible to be appraisers.

Findings from this exercise helped assess the AGREE II instrument itself, as a valid framework to:

- assess the quality of guidelines in the field of health protection and communicable disease control; and
- inform about a more suitable set of quality criteria to validate guidelines for health protection as well as guideline development methods for health protection.

To facilitate this exercise, we used an 'assessment table' where the group collated members' feedback with regards to each of the quality items/criteria the AGREE II instrument proposes. Feedback was focused on the use, significance, relevance and applicability of each of the quality criteria proposed by AGREE II and in relation to guidelines, or to guidelines development for health protection and public health.

4.5 Results/Findings

This section highlights the areas where the AGREE II instrument is likely to need some adaptation to make it suitable for the appraisal of communicable disease guidelines. However, it is acknowledged that the AGREE II instrument requires further testing with a broader range of communicable disease guidelines before more definitive conclusions and recommendations can be made.

Feedback and comments from this exercise are further expanded in the following four areas:

- the AGREE II terminology;
- on the instrument items;
- on the domains; and
- suggested additional criteria in relation to its application to health protection and communicable disease control.

4.5.1 AGREE II Terminology

The language of the AGREE II is appropriate to public health: based on this exercise, it appears that AGREE II works well for the purpose of evaluating communicable disease guidelines. Some in relation to terminology

suggested amendments would be to replace 'health question(s)' with 'key question(s)' in Item 2 and refer to different options for 'intervention' instead of 'management' in Item 16.

4.5.2 AGREE II Items

Most of the AGREE II items are applicable or partially applicable to communicable disease guidelines, at least in circumstances where there are no time and evidence constraints.

However, the following comments refer to the AGREE items/domains:

a) Domain: Scope and purpose (three criteria)

- **Items 1 and 2: The overall objective(s) of the guideline is (are) specifically described and the health question(s) covered by the guideline is (are) specifically described.** Requiring 'specific description' of objectives or questions does not actually mean the document meets these objectives or explores all possible answers to the proposed questions.
- **Item 3: The population (patients, public, etc.) to whom the guideline is meant to apply to is specifically described.** Reference to population targeted is covered by criteria 1 and 2.

b) Domain: Stakeholder involvement (three criteria)

- **Item 4: The guideline development group includes individuals from all the relevant professional groups.** This is an important issue in communicable disease, as the guideline development group are potentially guiding on decisions that affect whole populations without their say or explicit consent. This item could be extended so that there is a requirement to describe the process for determining which types of professionals are needed on the group. It could also cover how the members were selected/ recruited (e.g. open advert, recruitment policy, equal opportunities). There may be 'generic requirements' to involve some specific professionals – for example general practitioners are always in both ends of hospital admissions and play an important role in treatment for a number of infectious diseases/outbreaks.
- **Item 5: The views and preferences of the target population (patients, public, etc.) have been sought.** Taking account of the views and preferences of the target population is often difficult in public health/communicable disease guidelines: although representation from the public can be sought in the guideline development group, real public involvement is questionable. Referring to how public representation is recruited may be helpful to qualify this item. Once again, this item requires greater emphasis because of the potential for communicable disease recommendations to affect public freedoms etc.
- **Item 6: The guideline has been piloted among end users**. There is some overlap with item 1, in that the intended users of the guidance should be set out in the overall objective.

c) Domain: Rigour of development (eight criteria)

- **Item 7: Systematic methods were used to search for evidence.** Time pressures may make it difficult to conduct extensive/comprehensive searches, which might yield an unmanageable number of references to sift through. It might be quicker and ultimately more useful to identify the most relevant references through expert networks/key contacts, snowballing through references, etc. It would be useful to compare the results of these different approaches. In addition, some of the most useful evidence might not be found in peer reviewed journals or via the standard commercial databases. It might be in reports of outbreaks found on organisational websites or unpublished documents or via key contacts.
- **Item 8: The criteria for selecting the evidence are clearly described**. Details on the search strategy undertaken could be supplied (including inclusion/exclusion criteria, criteria for abstracts selection), however, this might not be possible under time and evidence pressures.
- **Item 9: The strengths and limitations of the body of evidence are described.** Individual studies should be quality appraised so that the guideline developers can determine how much they can rely on the findings (see Chapter 3 in this report for further details of critical appraisal tools for different study designs). This should extend to an overall assessment of the strength of the body of evidence (quantity, quality and consistency) and its applicability, so that the guideline developers can judge whether this is sufficient for making recommendations. Ideally, there should be a separate description of the guideline development group's 'consideration' of the evidence for developing recommendations (see also criteria 11 and 12).
- **Item 10 and 11: The methods for formulating the recommendations are clearly described and the health benefits, side effects and risks have been considered in formulating the recommendations.** A description of the methods for formulating the recommendations is often the least well covered in any guideline because of the complex nature of using groups to develop recommendations. Having said that, it should be possible to set out whether formal or informal methods were used and to describe the kinds of factors that the group was asked to take into consideration (see Chapter 5 in this report for further discussion of consensus methods). It should also be possible to summarise some of the guideline development group debates – this could be covered in the guidelines itself or in minutes of the meetings to be made publicly available. Details on how benefits, side effects and risks have been considered

is often not explicitly explained in guidelines, although risk assessment is expected and required in any public health/communicable disease guidelines. Implicitly, however, the 'considered judgement' phase (SIGN 50) within the guideline methodology encourages guideline development group members to consider risk assessment and the quality measure elements suggested on this criteria (see Annex 6 for further details). Explicit reference to this may be required, particularly in guidelines for public health/ communicable disease. It should be possible to set out the guideline development group's deliberations about the evidence in a (semi-)structured way, so that the reader can understand how the group arrived at its recommendations. The recommendations should also allow for different actions/options according to varying risks and benefits in different circumstances (see also criteria 9 and 12).

- **Item 12: There is an explicit link between the recommendations and the supporting evidence.** Making an explicit link between the recommendations and the supporting evidence is a crucial to the development of good quality guidance. This maybe the 'key criteria' to qualify guidelines as 'good' or 'less good'. The challenge remains when we consider guidelines in time and evidence constrained circumstances. In terms of quality appraising a guideline, another aspect to consider is how much 'a reader' assessing a guideline can afford in 'digging in' the evidence to consider that link between the recommendations and the evidence really exists.
- **Item 13: The guideline has been externally reviewed by experts prior to its publication.** Even with time pressures, it should be possible to have some external review before publication. It may be that some of the stakeholders are expert groups, so there is an opportunity through stakeholder consultation (and specific questions could be posed as part of this).
- **Item 14: A procedure for updating the guideline is provided.** It should be possible to comply with this item in any circumstances, indeed it is particularly important if the guidance has been developed rapidly and there was not sufficient time to gather evidence, involve experts and stakeholders in the original version.

d) Clarity of presentation (three criteria)

- **Item 15: The recommendations are specific and unambiguous.** Regardless of the time and evidence available, all recommendations should specify: Whose health will benefit (population)? Who should take action? What action(s)? What circumstances?
- **Item 16: The different options for management of the condition or health issue are clearly presented**. Describing management options is perhaps a challenging request in outbreaks management guidelines. This item has some overlap with item 15. There may be different recommendations or actions for different subpopulations and/or in different circumstances.
- **Item 17: Key recommendations are easily identifiable**. This item is only partially applicable, given that it may not be necessary or helpful to identify 'key' recommendations. Either all recommendations may be inter-related and necessary, or some are more 'key' for different audiences. Given that public health guidance can have very many users, it is often not appropriate to list any recommendations as key.

e) Applicability (four criteria)

- **Item 18: The guideline describes facilitators and barriers to its application.** Information on facilitators and barriers could be considered as part of initial or final drafting and they should be documented in the guidance or the minutes of the guideline development group. They could come from:
  - literature reviews of patient/professional views;
  - guideline development group's experience;
  - expert testimony;
  - surveys of current practice;
  - stakeholder comments on draft guidance;
  - stakeholders/expert;
  - piloting (or 'field-testing') of draft guidance.
- **Item 19: The guidelines provide advice and/or tools on how the recommendations can be put into practice.** Although these materials might be developed separately, the guidance should refer to where and when related tools and support might be found (e.g. a web address for further details).
- **Item 20: The potential resource implications of applying the recommendations have been considered.** The cost effectiveness of different 'options' or 'interventions' ought to be considered during the development of recommendations. This can be done by reviewing existing economic evaluations and/or economic modelling of interventions or 'scenarios'. However, the resource impact of a particular guidance can only be done when the recommendations have been finalised – hence it is often not included in the guideline document itself. Although, some of the resource implications may have been covered under 'barriers and facilitators' in item 18. Some areas of public health/communicable disease may be particularly challenge if cost effectiveness is required a priori.

- **Item 21: The guideline presents monitoring and/or auditing criteria.** This item is difficult to assess and depends on the nature of the public health issue and the types of recommendations. Many recommendations might be only partially implemented, depending on their complexity and appropriateness to the situation. Audit in turn is able to inform guideline reviews and further improve the implementation of specific recommendations.

f) Editorial independence (two criteria)

- **Item 22: The views of the funding body have not influenced the content of the guideline.** In the case of most public health guidance, the funder will be government or public sector agency. Given the potential 'political' decisions and consequences involved in some communicable disease interventions, it is important that the development of recommendations is genuinely independent and transparent.
- **Item 23: Competing interests of guideline development members have been recorded and addressed**. There should be a process for everyone participating in committee meetings, including committee members, experts and co-optees, members of the review teams and secretariat, to declare any financial, academic, personal or organisational interests and these should be on the public record.

4.5.3 Additional criteria for communicable disease guidelines

Although most of the AGREE II items seem to be relevant to communicable disease guidelines, the following additional criteria should also be considered:

- **Communicating the recommendations to patients, public and media**: given that communicable disease guidelines may have implications for civil liberties and/or an impact on population behaviour, it is important that they key messages are communicated in a clear and timely manner. While this could be incorporated into other items, it is suggested that it should be a stand-alone item, either within one of the existing domains (Clarity or Applicability) or in a new domain specific to communicable disease. Equality and diversity issues should be also considered (e.g. does the guidance avoid unlawful discrimination?).
- **Consideration of delivery structures and mechanisms,** given the integrated nature of public health systems, networks, professionals. This Item could be added to the Applicability domain.
- **Consideration of legal and regulatory frameworks**, national policy, statutory guidance, etc. e.g. guidelines on management of *Legionella* outbreaks.
- **Role of ethics** and social value judgements (equity and equality) and potential to cause or prevent harm to individuals, e.g. through onward transmission. Ethics is being advocated as an element of decision-making in public health/communicable disease interventions/outbreak management. Evidence and theory relating to 'effectiveness' should serve the full range of ethical principles, not just 'do good'. Other principles: equity, respect, empowerment, sustainability, social responsibility, participation, openness, accountability. Reflection on the use of ethical analysis in public health practice and on the development of public health policies is becoming more important in pandemic and infectious disease control scenarios in which professional roles and responsibilities generate ethically complex situations (e.g. powers and duties of public health professionals, ethical issues related to preparing and responding to outbreaks).
- **Health economics**: public health guidelines may often be costly to implement. The direct costs of implementing guidelines should always be presented, when possible alternative costs and societal gains and costs should be calculated to facilitate the final decisions and judgements to be done by the policymakers. The use of generic comparable estimates like 'Quality Adjusted Life Years' are recommended.
- **Trade-off between harms and benefits.** Where possible, HP guidelines should assess any potential negative effects and whether these are offset by the anticipated benefits.

4.5.4 AGREE II domains

The domain concerned with 'Rigour of development' is the largest domain (it has twice or more the number of criteria of the other domains) and arguably it is one of the most important domains, given it goes to the heart of evidence-based medicine/methodology. There appears to be some overlap between several items (9–12), both in terms of their focus (what they are trying to assess) and in terms of where to look and how to rate. Nevertheless, it should be possible for public health guidelines to score highly on most of them, even under time and evidence pressures (exceptions noted above, e.g. item 7). Even then, it might be possible for guideline developers to maintain databases of pre-prepared search strategies, bibliographic lists of public health references and pre-appraise studies that are likely to be needed in future guidelines.

The 'Applicability' domain presents a challenge for most clinical guidelines, even more so for public health guidelines. Nevertheless, most of the criteria apply; indeed this domain could be extended to include some additional criteria for communicable disease guidelines (see below).

In summary, the AGREE II domains are appropriate to communicable disease guidelines and, in principle, even under circumstances of time and evidence constraints. However, some consideration should be given to introducing an additional domain(s) if the suggestions for further criteria set out above cannot be incorporated into

the existing domains. In addition, a 'short' version of the instrument based on a selection of domains, could be used to evaluate or develop guidance under time and evidence constraints (this will be introduced in section 4.6.1).

4.6 Discussion and conclusion

The exercise carried out by this *working group* proved to be a rich source of material, providing insight into harmonising and standardising guideline development and guideline quality assessments. The use of the AGREE II instrument has helped systematically analyse the structures and working methods of guideline programmes as they apply to health protection and communicable disease control.

The credibility provided by the use of AGREE II instrument is widely acknowledged. The need for harmonising and standardising guideline development and quality assessment is considered one of the most important reasons that prompted the establishment of international organisations, such as the Appraisal of Guidelines for Research and Evaluation (AGREE) Research and later Guidelines International Network (G-I-N).

However, it remains to be determined which individual criteria are essential in all circumstances for public health (communicable disease control) guidelines and which could be sacrificed if evidence and time are an issue. Or at least, what we might accept as substitute rating criteria (e.g. in place of systematic searches). Further, could 'quality level bars' be defined for each criteria, depending on time and evidence constraints circumstances? E.g. although patients/public involvement may be desirable, in a situation of lack of time this might not be feasible.

4.6.1 Guideline Evaluation Tool (GET5) – A shorter and adapted version of the AGREE II

For circumstances of time and evidence constraints, we suggest the use of a shorter and adapted version of the AGREE II instrument that has been labelled as GET5 (Guideline Evaluation Tool) – see Appendix 6. GET5 has resulted from preliminary work carried out in Health Protection Scotland (HPS) and trialled in-house for rapid assessment of guideline documents. The use of GET5 is suggested in this report to rapidly appraise guidelines for communicable disease control:

- after being rapidly produced in time and evidence constraint circumstances;
- while considering various guidelines that could be part of the body of evidence considered to inform new guideline being developed;
- while reviewing information resources available by your organisation (either routinely or in the circumstances of an incident/outbreak).

4.6.2 Recommendations for further research/further developments

It is acknowledged that the AGREE II instrument and the GET5 tool require further testing with a broader range of communicable disease guidelines before more definitive conclusions and recommendations can be made. This might include further consideration of:

- clarity and consistency in the terminology used in appraisal exercises and in Guidance for public health/communicable disease: e.g. users, stakeholders, validation, quality, expert review, etc.
- more 'universal' use of a more applicable appraisal system for public health/communicable disease guidelines.

There is need for a further collaboration with the AGREE foundation to validate (and extend) the use of the AGREE II instrument for public health guidance.

In addition, it would be useful to consider developing and validating (either with the AGREE collaboration or as a health protection/communicable disease control community of experts) an appraisal tool to assess the content of guidelines, in particular the robustness of the link between the recommendations and the underlying evidence.

The AGREE III (the study), which aims to create tools to facilitate and reliably identify the specific features used by guideline development panels in making judgments during the development of their recommendations, should integrate public health considerations.

Finally, the need of establishing a register of public health guidances produced by national agencies, potentially via G-I-N, should be considered. This register should also contain mechanisms for sharing relevant evidence reviews on public health topics with organisations worldwide.

4.7 References

[1] Woolf SH, Grol R, Hutchinson A, Eccles M, Grimshaw J. Clinical guidelines: potential benefits, limitations, and harms of clinical guidelines. BMJ. 1999 Feb 20;318(7182):527-30.

[2] Field MJ, Lohr KN. Guidelines for Clinical Practice: From Development to Use. Washington DC: National Academy Press; 1992.

[3] Sekimoto M, Imanaka Y, Kitano N, Ishizaki T, Takahashi O. Why are physicians not persuaded by scientific evidence? A grounded theory interview study. BMC Health Serv Res. 2006 Jul 27;6:92.

[4] Codish S, Shiffman RN. A model of ambiguity and vagueness in clinical practice guideline recommendations. AMIA Annu Symp Proc. 2005:146-50.

[5] Shiffman RN, Shekelle P, Overhage JM, Slutsky J, Grimshaw J, Deshpande AM. Standardized reporting of clinical practice guidelines: a proposal from the Conference on Guideline Standardization. Ann Intern Med. 2003 Sep 16;139(6):493-8.

[6] Harpole LH, Kelley MJ, Schreiber G, Toloza EM, Kolimaga J, McCrory DC. Assessment of the scope and quality of clinical practice guidelines in lung cancer. Chest. 2003 Jan;123(1 Suppl):7S-20S.

[7] Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project. Qual Saf Health Care. 2003 Feb;12(1):18-23.

[8] Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. BMJ. 1996 Jan 13;312(7023):71-2.

[9] Field MJ, Lohr KN. Clinical Practice Guidelines: Directions for a New Program. Washington, DC: National Academy Press; 1990.

[10] Wollersheim H, Burgers J, Grol R. Clinical guidelines to improve patient care. Neth J Med. 2005 Jun;63(6):188-92.

[11] Eccles M, Clapp Z, Grimshaw J, Adams PC, Higgins B, Purves I, et al. North of England evidence based guidelines development project: methods of guideline development. BMJ [serial on the Internet]. 1996; 312(7033): Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2350465/pdf/bmj00534-0046.pdf.

[12] Sheldon TA, Cullum N, Dawson D, Lankshear A, Lowson K, Watt I, et al. What's the evidence that NICE guidance has been implemented? Results from a national evaluation using time series analysis, audit of patients' notes, and interviews. BMJ. 2004 Oct 30;329(7473):999.

[13] Burgers JS, Grol RP, Zaat JO, Spies TH, van der Bij AK, Mokkink HG. Characteristics of effective clinical guidelines for general practice. Br J Gen Pract. 2003 Jan;53(486):15-9.

[14] Grol R, Dalhuijsen J, Thomas S, Veld C, Rutten G, Mokkink H. Attributes of clinical guidelines that influence use of guidelines in general practice: observational study. BMJ [serial on the Internet]. 1998; 317(7162): Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC31096/pdf/858.pdf.

[15] Foy R, MacLennan G, Grimshaw J, Penney G, Campbell M, Grol R. Attributes of clinical recommendations that influence change in practice following audit and feedback. J Clin Epidemiol. 2002 Jul;55(7):717-22.

[16] University of York. NHS Centre for Reviews & Dissemination. Getting Evidence into Practice. Plymouth, Great Britain: University of York, NHS Centre for Reviews & Dissemination; 1999.

[17] Burgers JS, Bailey JV, Klazinga NS, Van Der Bij AK, Grol R, Feder G. Inside guidelines: comparative analysis of recommendations and evidence in diabetes guidelines from 13 countries. Diabetes Care. 2002 Nov;25(11):1933-9.

[18] Burgers JS, Cluzeau FA, Hanna SE, Hunt C, Grol R. Characteristics of high-quality guidelines: evaluation of 86 clinical guidelines developed in ten European countries and Canada. Int J Technol Assess Health Care. 2003 Winter;19(1):148-57.

[19] Michie S, Johnston M. Changing clinical behaviour by making guidelines specific. BMJ. 2004 Feb 7;328(7435):343-5.

[20] Graham ID, Calder LA, Hebert PC, Carter AO, Tetroe JM. A comparison of clinical practice guideline appraisal instruments. Int J Technol Assess Health Care. 2000 Autumn;16(4):1024-38.

[21] Cluzeau FA, Littlejohns P, Grimshaw JM, Feder G, Moran SE. Development and application of a generic methodology to assess the quality of clinical guidelines. Int J Qual Health Care. 1999 Feb;11(1):21-8.

[22] Vlayen J, Aertgeerts B, Hannes K, Sermeus W, Ramaekers D. A systematic review of appraisal tools for clinical practice guidelines: multiple similarities and one common deficit. Int J Qual Health Care. 2005 Jun;17(3):235-42.

[23] Scottish Intercollegiate Guidelines Network (SIGN). Clinical guidelines: criteria for appraisal for national use. Edinburgh: Royal College of Physicians; 1995.

[24] Helou A, Ollenschlager G. [Goals, possibilities and limits of quality evaluation of guidelines. A background report on the user manual of the 'Methodological Quality of Guidelines' check list]. Z Arztl Fortbild Qualitatssich. 1998 Jun;92(5):361-5.

[25] Oxman AD, Schunemann HJ, Fretheim A. Improving the use of research evidence in guideline development: 16. Evaluation. Health Res Policy Syst [serial on the Internet]. 2006; 4: Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1702533/pdf/1478-4505-4-28.pdf.

[26] Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, et al. AGREE II: advancing guideline development, reporting and evaluation in health care. J Clin Epidemiol. 2010 Dec;63(12):1308-11.

[27] Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, et al. Development of the AGREE II, part 1: performance, usefulness and areas for improvement. CMAJ [serial on the Internet]. 2010; 182(10): Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2900328/pdf/1821045.pdf.

[28] The Agree Next Steps Consortium. AGREE II instrument. [s.l.]: The AGREE Research Trust; 2009. Available from: www.agreetrust.org/index.aspx?o=1397.

[29] National Institute for Health and Clinical Excellence. Methods for Development of NICE Public Health Guidance (2nd ed.). London: NICE; 2009.

[30] National Institute for Health and Clinical Excellence. The NICE public health guidance development process: an overview for stakeholders, including public health practitioners, policy makers and the public (2nd edition). London: NICE; 2009. Available from: http://www.nice.org.uk/media/F19/70/PHProcessGuide2009.pdf.

[31] Kelly M, Morgan A, Ellis S, Younger T, Huntley J, Swann C. Evidence based public health: A review of the experience of the National Institute of Health and Clinical Excellence (NICE) of developing public health guidance in England. Soc Sci Med. 2010 Sep;71(6):1056-62.

[32] Health Protection Network. Guideline for the control of measles incidents and outbreaks in Scotland. Health Protection Network Scottish Guidance 4. Glasgow: Health Protection Scotland; 2010. Available from: http://www.documents.hps.scot.nhs.uk/about-hps/hpn/measles-guidelines.pdf.

[33] Health Protection Agency. Guidelines for the investigation, control and prevention of spread of group A streptococcal (GAS) infection in acute healthcare settings in the United Kingdom. [updated 9 August 2010]; Available from: http://www.hpa.org.uk/web/HPAweb&HPAwebStandard/HPAweb_C/1272032560620.

5 EBM methods for public health – the use of consensus methods

5.1 Background and aim

Public health institutions often deal with the problem of trying to make decisions in situations where there is insufficient (or even contradictory) information. While evidence-based methods have been developed to evaluate existing data and study findings, these methods imply that information is available in an appropriate form. Consensus methods provide another means of generating advice. These methods are capable to use a different range of information when compared to evidence-based methods. They provide an approach of harnessing the expertise of appropriate experts to enable decisions where published information is inadequate or non-existent.

If scientific evidence is available and evidence-based methods are applied, consensus methods are also a valuable tool to formalise the involvement of appropriate experts in the interpretation of this evidence in order to improve understanding and practicability of public health guidance.

The aim of this section of the report is to review the strengths and weaknesses of different consensus methods in the field of public health, and to evaluate them with respect to the specific needs of rapid public health advice.

5.2 Methods

The members of the expert group developed this chapter through interactive discussions during three face to face meetings held in April 2009, September 2010 and December 2010, respectively, followed by circulation of draft report for comments and revisions. The first meeting held in plenary defined the group mandate and listed the questions to be addressed by the group, the second meeting revised and reformulated these questions as well as elaborated draft answers, which were further discussed and expanded in the plenary group at the last meeting. Additional input was offered as written comments by the other members of the group. The subgroup moderator integrated these comments into the revised draft, which was reviewed and approved by the subgroup members.

5.3 Definition

Consensus: *group agreement on opinions and judgments, reached as a whole*; derived from the Latin *consentire*, literally 'to feel with' and the underlying meaning, 'to feel the same'. A consensus is a strongly held opinion, where the majority feel strongly enough about something to make the stance together [1].

The role of consensus methods in health services research and policymaking is to determine the extent to which experts or lay people agree about a given issue [2].

5.4 What is the place of consensus methods in the development of scientific guidance for public health?

The development of expert appreciation of the best available scientific evidence, once it is summarised by systematic review and meta-analysis of scientific studies, can improve the balance of its subjective interpretation and its translation into 'actionable' options. In addition, obtaining and summarising the consensus expert opinion and collective wisdom of health practitioners should allow pragmatic input from experience and ensure professional relevance. Lastly, the involvement of other stakeholders, including lay persons concerned by the health issue under review, in the development of consensus support to the guidance or its formulation may enhance its uptake by service providers and acceptability to users.

The type of consensus development process that is desirable and feasible will depend on a number of elements of public health guidance, such as its geographical and societal scope (local, national, international), its focus and operational level (risk assessment, risk management, quality of care improvement, health education, health technology assessment), target audience and users (public, healthcare practitioners, policymakers), quality of available evidence, time pressure and resource constraints.

It is important to distinguish three distinct situations at which consensus can be sought on public health scientific guidance.

1. The first situation relates to the development of broad political and societal consensus about the **generic process** used for public health guidance development. At national level, it is desirable that the guidance development procedures are explicitly defined and include content validation through a public consultation

phase. This is illustrated, for instance, by the Swedish experience on public consultation on drafts of national health board guidance. This process, which involves two rounds of consultation using web-based communication and other media, systematically allows expression of lay opinion and input from end users that enhance social acceptability of guidance outputs. It also guarantees a high level of public accountability, democratic transparency and consistency of national guidelines. At international public health agency level, such as ECDC, broad stakeholder endorsement of generic scientific advice development process may be achieved by using a guidance development protocol approved by the competent bodies in EU Member States.

2. The second situation refers to the use of **consensus methods** by members of the **guidance development group** as part of their draft guidance elaboration. This may cover several goals: to reconcile or delineate differences in expert views on summarising and appraising the literature; to address gaps in available scientific evidence by unpacking expert experience and knowledge of grey literature, and harnessing their a priori knowledge and understanding of biological determinants of communicable disease risk.

3. The third situation is about use of 'consensus' or 'disagreement' measurement methods at the later stage of **draft guidance output validation**. This may include input from diverse approaches: scientific external peer review and stakeholder consultation, such as healthcare practitioners, members of target populations or the general public. It is obviously cumbersome and impractical to involve external experts and other stakeholders at every step of the guidance development, but valuable to do so at the final stage, whenever time and resources so allow.

5.5 Improving transparency of public health decision-making

The overarching goal is to support a transparent and systematic process of judgment of available knowledge, with the purpose of taking well-informed decisions about disease prevention and healthcare, even under scientific uncertainties and time pressure.

Whereas the transparency and quality of judging scientific evidence in a clinical setting has improved in recent years through the use of quality appraisal tools for studies on diagnosis, treatment, prognosis and harm, fewer systematic tools and procedures have been developed for the 'softer' elements of a public health decision-making process. The considered opinions and judgments from stakeholders, content experts, patient representatives, populations at risk, political and administrative decision-makers should be valued, graded and expressed with a common understanding of level of confidence, degrees of uncertainties and the acceptability of a potential decision The challenge is to develop a framework where the best EBM methods on grading of scientific evidence is merged with an explicit and transparent process of developing the considered judgments that are essential parts of the final decision-making process.

Heterogeneity among studies is often considered as a main problem when collating evidence from different studies of intervention effects. When trying to standardise and quantify the elements and the judgments of a final public health decision, heterogeneity will be a huge problem. Methodologists might say it is impossible to solve, but possible to describe. One problem is that in the process of public health decision-making the considerations about ethics, law, economy, epidemiological, social and political context, patient and population's preferences are dealt with on a daily basis, and the values of the different elements are compared, weighted and evaluated often in an unsystematic, implicit and non-transparent way.

It has been claimed that, in areas of public health policy characterised by major uncertainty on risk magnitude and limited or conflicting evidence on management options, it is preferable to refrain from 'forcing' 'authoritative' consensus risk assessment based on shaky grounds. Instead, it is proposed to provide 'plural and conditional advice', by presenting alternative expert interpretations of the evidence and recommended options with explicit reference to their assumptions, values and intentions [3].

5.6 The role of experts

These considerations and judgments have to be done at several stages of a scientific advice process and expert input is needed to take well-informed decisions. An expert has often been considered as a person with extensive in-depth knowledge and research experience in his field, and as such subject to a potential academic bias as well, towards other views and other experts. Due to this personal dominance bias, a deep scepticism towards the capacity of content experts to perform independent and balanced judgment of evidence has often been expressed by the EBM methodologists.

It is generally accepted that experts are valuable in understanding and evaluating the core content matters of a health problem, and such experts are often easy to identify by their academic and research credentials. Less attention has been paid to ways of identifying the best experts on social context, ethics, patient values, practical feasibility, political procedures and processes of different countries at a local, regional, national and international level.

An expert who has a lot of knowledge and experience can give very valuable input to a decision-making process when time to do extensive searches for evidence or to get information through hearings is limited. The less time there is, the more you need the experts. To improve transparency, it is desirable that experts spell out explicitly in their statements what is based on knowledge derived from their own experience or research, sources of information in the 'grey literature' or personal opinion. Furthermore, they should report possible conflicts of interests, academic as well as financial.

In decision-making we will always want to synthesise the best available evidence or knowledge about as many of the relevant variable factors as possible in the process. The challenge is how to weigh and possibly quantify the different factors against each other. Here again, expert judgment will play a major role in deciding what is the weight of the evidence about harm versus benefit, about risks versus consequences, about costs of interventions versus alternative costs or about ethics versus utility.

5.7 Selecting the appropriate experts

The **composition** of the guidance development group needs to be tailored to each topic. It should be based on explicit a priori **criteria defining the type of expertise, experience or perspective to be included**, and describing how each participant meets the specifications, such as:

- health professional and scientific expertise: content expertise, methodology, public health, epidemiology, sociology, health services, quality of care, health economy, etc.
- stakeholders: patient organisations, health authorities, healthcare service providers, service payers, health industry, at-risk population groups, general public.

For international guidance, an additional dimension to be included in the criteria for appointing experts is the range of countries, reflecting diversity of target audiences with respect to socio-economic context, healthcare and public health systems and cultural values.

5.8 Conflicts of interest

To ensure public appraisal of the independence and possible conflicts of interest that may have biased the judgment of experts, public declaration of their financial ties, academic interests, and institutional affiliations should be the rule and incompatibility criteria, if applicable, be reported explicitly.

5.9 Use of consensus methods by guidance development group

Consensus within the multidisciplinary **guidance development group** should be reached at the following **critical and subjective steps** of the development of evidence-based guidance:

1. defining the key questions, selecting the relevant data sources and type of evidence to be retrieved;
2. determining which outcomes are critical to the grading of evidence and, therefore, informing strength of recommendation;
3. translating the summary of scientific evidence into draft recommendations;
4. draft revision according to external comments and ratings from external review and consultations; and
5. endorsing the final recommendations.

The **facilitator** and chair of the guidance development group should be selected for his/her communication and group management abilities as well as conflict resolution skills. He/she should preferably not have strong academic bias nor hold direct responsibility for implementation of guidance outputs. However, he/she should possess sufficient content knowledge and authority to lead discussions constructively and resolve, or at least clarify, disagreement where relevant. Otherwise, back-up content experts should be involved in co-moderation. Formal training in facilitation process is increasingly considered desirable and may be provided to guidance group moderators.

5.10 The use of consensus methods for guidance development

Formal **consensus building and assessment methods**, such as the Delphi method, Nominal Group Techniques, and others, all have their specific strengths and weaknesses [2].

In the Delphi method, a first questionnaire survey allows the group participants to privately express their opinion on a particular question. These opinions are then summarised and organised in a limited set of statements, which are then circulated to allow participants to rank their agreement with these statements in the questionnaire. The results are summarised and circulated to all participants with a repeat version of the questionnaire for a second round of rankings. The final rankings are summarised and assessed for degree of consensus and the participants receive feedback.

In the nominal group technique, a group of up to 12 experts participate in a structured meeting where the facilitator collects each participant view on a specific question for collation of group opinions. These are then discussed with the group to evaluate, clarify and re-organise proposals. In a first rating round, each participant privately ranks each statement. The overall ranking is tabulated and discussed. A final round of ranking is performed, tabulated and communicated to the participants.

Variations and combinations of these methods have been widely used in different fields to gather group consensus and help with guidelines development [4,5]. Consensus methods have a number of common features. First, they allow equal opportunity for each expert or group participant to express his or her own judgment by using anonymous ratings to avoid dominance by the most vocal members of the group. Second, they allow for group interaction and iterative judgment processes by using several rounds of appropriateness ratings and reformulation of statements. Third, they provide a controlled and systematic feedback to each participant, indicating how his previous response compares to the distribution of opinion across the group. Finally, they produce a statistical summary estimate of the group adherence to the draft and final statement.

No particular method was considered by the group to be generally superior in helping to facilitate discussion, optimally pool the collective wisdom and range of expert views, and quantify the degree of consistency of group judgment. It was felt that a combination of formal individual rating and interactive group discussion was generally desirable. The choice of consensus methods and length of process will depend on the time and resource constraints and desirable range of stakeholder involvement. In any case, the criteria used for scoring appropriateness and thresholds of agreement for final recommendations should be agreed by the group before appropriateness ratings are performed.

As a rule of thumb, it was proposed that the following options may be considered:

- For **urgent (2–3 days) risk assessment**, a single rapid round of collegial peer-review of a working draft document is often the best expert validation that can be realistically achieved.
- For **semi-urgent (2–3 months) risk assessment and public health guidance**, at least one round of remote appropriateness ratings of narrative summaries of evidence and derived draft recommendations by a Delphi type method would be desirable.
- For **non-urgent (6 months to 2 years) risk assessment and public health guidance,** it was considered preferable to use a minimum of three rounds of internal iterative elicitation of opinion for producing draft recommendations: 1) a first round of decentralised appropriateness ratings using electronic questionnaires (Delphi rounds or similar method); 2) a face-to-face meeting of the experts to debate their first Delphi results and amend the draft; 3) a second round of remote Delphi ratings of the revised draft. Furthermore, it is desirable to validate the resulting draft guidance by consulting users and other stakeholders who may be affected by application of the guidance, to enhance the 'buy-in' by society at large.

Irrespective of time constraints and method(s) chosen to gather expert knowledge and consensus opinion, **transparency of the guidance decision-making process** should be ensured. Formal consensus methods can help trace the processes and the discussions that led to the conclusions by explicitly reporting the link of recommendation to evidence, criteria applied to elicit considered judgement of feasibility, quantifying the extent of expert agreement and describing areas and reasons for disagreement.

5.11 References

[1] V2 Vocabulary Building Dictionary. Available from: http://vocabulary-vocabulary.com/

[2] Jones J, Hunter D. Consensus methods for medical and health services research. BMJ. 1995 Aug 5;311(7001):376-80.

[3] Stirling A. Keep it complex. Nature. 2010 Dec 23;468(7327):1029-31.

[4] Raine R, Sanderson C, Black N. Developing clinical guidelines: a challenge to current methods. BMJ. 2005 Sep 17;331(7517):631-3.

[5] Fretheim A, Schunemann HJ, Oxman AD. Improving the use of research evidence in guidelines development: 5. Group processes. Health Res Policy Syst [serial on the Internet]. 2006; 4: Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1702534/pdf/1478-4505-4-17.pdf.

Annex 1 – The information specialist as part of an interdisciplinary team in the evidence-based process

The need for the involvement of an information specialist can be identified at different stages of the evidence-based process. The unique professional skills of the information specialist cover expert searching, organisational expertise and document retrieval. They should take an active part in an interdisciplinary team working with an evidence-based method.

Searching for the evidence

Information specialists play an important role when looking for evidence. Information specialists and experts in the subject field should have a near, collaborative and interactive work in the first stage of defining the concepts for the search strategies. They both will have to invest time together checking the results and agreeing in the redesign of the searches and/or incorporating new concepts according to the relevant questions.

The information specialist has the expertise in choosing the right sources of information, such as bibliographic databases, websites and/or grey literature sources. Standard 3.1, produced by the Institute of Medicine of the National Academies, describes in detail the steps of the work of an information specialist or librarian for conducting a comprehensive systematic review for evidence [1].

The use of Boolean and proximity operators and truncation options are common in database search strategies. The information specialist is trained to identify the best way to use these options together with keywords, thesaurus terms in the database and indexes. The retrieval of information becomes more complex when all these elements vary or they have a different usage among the different sources of information. Hence, information specialists' expertise becomes necessary not only for their skills in searching information but also for validating the source of information and the knowledge they contain in aspects like subject, scope and time coverage [2]. The Medical Library Association has developed a policy statement for the definition of expert searching and the skills of health science librarians (named information specialists in this document) for providing expert searching. It also identifies those areas of expertise that healthcare and biomedical professionals do not have and information specialists are trained for [3].

Database-storing references and access to full text

The organisational skills of the information specialist are applied to control, organise and maintain a database that stores all the references retrieved from all the sources of information [4]. At this point, a citation management software becomes a very useful tool because, among other features, it provides the advantage of combining all references from all the different sources consulted and eliminates duplicated references to ensure the consistency of the database created.

After the screening process and the selection of the relevant references by the experts, the information specialist locates and provides the full-text articles. If the references quoted are unavailable among the collection of the library or information centre, the information specialist locates them and obtains their full text via the Inter Library Loan service. This service allows requesting those references unavailable in the collection to other libraries or information centres.

Final report contribution

In the final report, the information specialist contributes with the explanation of the search methodology and the actual search strategies used during the retrieval of the information [2]. The search methodology is often located in the methodology section of the report. In this section, the information specialist explains:

- the sources of information used;
- the inclusion and exclusion criteria;
- the limits used;
- the dates when the information was retrieved; and
- any other relevant information related to the way the searches were performed.

The detailed search strategies for the different questions and sources involved could be included in the report as an annex, where the full search strings are stated, including the thesaurus terms, the keywords with the fields

applied (if applicable), the limits employed, and the combination of the different sets of concepts with Boolean operators.

Time constrains

In the public health domain the value of expert searching has become a key factor for providing timely and quality information for evidenced-based practice. Unfortunately, this fact is not properly considered due to lack time or resources (e.g. technology, the library or information centre is unavailable in the organisation or proper funds are insufficient) [5].

The timing is often an issue when providing a fast evidence-based answer due to political or media pressures. There is a responsibility to provide in the answers reliable information to policymakers. Hence there is a need to be up-to-date on the topic by receiving alerts instead of starting from the scratch the whole process. Among the services provided by libraries and information centres there is the Selective Dissemination of Information (SDI) service, which allows receiving automatically (often via email) the latest releases on a topic. This can be very useful to keep experts in the area updated.

Experts can contact information specialists in their organisations to set up an incoming information system in one specific topic that, among other elements, could be based on:

- the design of a search strategy in bibliographic databases, which will allow receiving recently published references that will match the search strategy saved. Bibliographic databases may also allow creating citation alerts;
- the selection of relevant organisations and subscription to their e-bulletins or RSS feeds. Organisations may have distribution lists for receiving the list of their new released publications;
- the choice of relevant scientific journals in the area and subscribe to their table of contents.

References

[1] Institute of Medicine of the National Academies. Finding what works in health care: standards for systematic reviews: standards March 2011. Available from: http://www.iom.edu/~/media/Files/Report%20Files/2011/Finding-What-Works-in-Health-Care-Standards-for-Systematic-Reviews/Standards%20for%20Systematic%20Review%202010%20Insert.pdf.

[2] McGowan J, Sampson M. Systematic reviews need systematic searchers. J Med Libr Assoc [serial on the Internet]. 2005; 93(1). Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC545125/pdf/i0025-7338-093-01-0074.pdf.

[3] Medical Library Association. Medical Library Association policy statement: role of expert searching in health sciences libraries. [Web page] Chicago, IL: The Association; 2003 [Consulted: 28/01/2011]. Available from: http://www.mlanet.org/resources/expert_search/policy_expert_search.html.

[4] Harris MR. The librarian's roles in the systematic review process: a case study. J Med Libr Assoc [serial on the Internet]. 2005; 93(1). Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC545126/pdf/i0025-7338-093-01-0081.pdf.

[5] Alpi KM. Expert searching in public health. J Med Libr Assoc [serial on the Internet]. 2005; 93(1). Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC545128/pdf/i0025-7338-093-01-0097.pdf.

Annex 2 – Tools for stages plus examples for modification / adaptation for use in rapid risk assessments

Stage	Tools	Examples
Stage 0: Preparatory	Ongoing alerting and verification systems	Early Warning and Response System (EWRS)
	Up-to-date systemic reviews and summaries of evidence	Cochrane reviews (http://www2.cochrane.org/reviews/) Attract (http://www.attract.wales.nhs.uk/)
	Quality standards for surveillance and field investigation and reporting	Adapt the STROBE and ORION guidelines for use http://www.strobe-statement.org/ http://www.oxfordjournals.org/our_journals/jac/press_releases/freepdf/dkm055.pdf Orion Statement
Stage 1: Incident verification	Protocol for verification of incidents with criteria on incident definition and alert levels	ECDC Epidemic Intelligence tutorial
Stage 2: Assessment of risk	Protocol for rapid searching and appraisal of peer reviewed and grey literature	to be developed or modified from existing search methods to fit purpose
	Protocol for communication between experts internationally	to be developed
	International database on incidents and reports	to be developed
	Protocol for sharing surveillance data internationally in a rapid fashion	e.g. HPS reporting influenza surveillance data on a daily basis to IMMFORM – DH web-based surveillance system
	Risk assessment framework for syntheses of evidence in relation to public health questions	version of SBAR – see Annex 3 http://www.documents.hps.scot.nhs.uk/hai/infection-control/toolkits/hiiat-2009-12.pdf http://www.documents.hps.scot.nhs.uk/hai/infection-control/toolkits/hiia-sop-2009-12.pdf http://www.institute.nhs.uk/quality_and_service_improvement_tools/quality_and_service_improvement_tools/sbar_-_situation_-_background_-_assessment_-_recommendation.html
Stage 3: Developing advice	Guidance on developing advice including strength of evidence and degrees certainty of effectiveness of interventions (explicit reference to advice for other sources)	Adapted and modified tools and methodology to be described
Stage 4: Implementation	Check list of key points to address in risk communication, including risk perception predictor, risk statement screening, media briefing template	To be developed
Stage 5: Monitoring and evaluation	Real time audit	To be developed

Annex 3 – Situation Background Assessment Recommendation (SBAR)

Originally used in the military and aviation industries, Situation Background Assessment Recommendation (SBAR) was further developed for healthcare and has become commonly used as a tool for multidisciplinary work in clinical care to aid communication and clarity. It has been promoted as part of patient safety initiatives. (See [1] for more information on the use within the NHS in the UK.)

SBAR has been used within the field of health protection and, specifically, in dealing with healthcare-associated infections and outbreaks of infection in hospitals in Scotland. It has become part of an outbreak management process to describe recommendations for action at different stages of the outbreak, and to clearly articulate the evidence underpinning the recommendations and, subsequently, the reason for changing the recommendations as the situation evolved and evidence changed. It helps to provide clarity and transparency to the process of outbreak management.

This method of communication was used to produce quick clear assessments on a couple of recommendations for action, which were produced as part of HPS response to the A(H1N1) influenza pandemic. The use of the SBAR tool is useful particularly in situations where specific questions need to be answered under time pressure. It enables the process to be detailed with the recommendations and would act as an audit trail for the process and decisions made.

Below is a modified version of the SBAR tool for a public health focus.

Question e.g. Is there a risk... to... from... influenza A(H1N1)v infection?	
Situation	Concern has been expressed due to reports of...
Background	**Threat assessment** Results will inform the level of risk Influenza is generally known to be a risk/or not... Include known evidence and expert opinion References – rapid literature search (& how? Pubmed, etc.) Risk assessment – see algorithm produced by ECDC as example? Evidence – or induced evidence – i.e. seasonal flu to H1N1, etc. Experts consulted
Assessment	**Summarise the known risks based on evidence and expert opinion** Pathogen Transmission At-risk populations Prevention/interventions Prophylaxis Treatment Country at risk or if evidence not available specifically on, e.g. a new infection, how the assessment is made based on existing knowledge and with the input of experts acknowledged Include the limitations – e.g. limited evidence – or based on different strain of infection Add information on whether this recommendation is taking a 'precautionary approach' while the evidence on the infection emerges? Acknowledge that this assessment may change as new evidence becomes available, e.g. based on the epidemiology of the new infection
Recommendation	**Include recommendation for action based on the rapid review of the available evidence and input of expert opinion**

Reference

[1] Institute for Innovation and Improvement. SBAR – situation-background-assessment-recommendation. Available from: http://www.institute.nhs.uk/quality_and_service_improvement_tools/quality_and_service_improvement_tools/sbar_-_situation_-_background_-_assessment_-_recommendation.html.

Annex 4 – HPA report on a methodology for rapid risk assessments (summary)

The following text is the summary from a report commissioned by ECDC to the Health Protection Agency (UK) on a methodology for rapid risk assessments.

This guidance develops a methodology for rapid risk assessments undertaken in the initial stages of an event or incident of potential public health concern. It describes an operational tool to facilitate rapid risk assessments for communicable disease incidents at both Member State and European level. The tool comprises an information table and risk ranking algorithm(s) to give an estimate of risk posed by a threat. The risk to a population from a communicable disease is dependent on the likelihood of transmission in the population (probability) and the severity of disease (impact). The probability of the incident developing or the impact if it does, are based on both the nature of the infectious agent and details of the incident. This may be further influenced by context or the broad environment in which the incident occurs, including political, public, media interest and perception of threat, and the acceptance of risk may vary between countries and cultures.

Rapid risk assessment is a core part of public health response and thus widely undertaken by public health professionals. Formal systems used to grade evidence and recommendations, such as the systematic methods used in evidence-based medicine (EBM) rely on published research evidence and studies are graded according to design and susceptibility to bias. However, as time and evidence are limited, rapid risk assessments may need to rely at least in part on specialist expert knowledge and these formal systems are not directly applicable. However, it is an aim to apply the same principles of transparency, explicitness and reproducibility even when doing a rapid risk assessment.

In rapid risk assessment for most infectious disease threats, only observational data is available and often this is the only possible obtainable sources of information. Expert knowledge is as well important when there is lack of time and limited evidence. In such cases it is important to 'unpack' the expert knowledge and to distinguish between knowledge based on good research and systematic gathered experience and opinion-based knowledge. As far as possible, attempts should be made to assess the quality of the evidence, based on the source, design and quality of each study or piece of information. Uncertainties should be identified and clearly documented and communicated and the assessment updated in light of new evidence over time.

A rapid risk assessment includes the approach to, and tools required, at each stage of the process: stage 0 is the preparation stage; stage 1 is the collection of event information; stage 2 is the literature search and systematic collection of information; stage 3, the extraction of evidence; stage 4, appraisal of the evidence; and stage 5, estimation of risk. Transparency and sharing of information is essential at every stage. The document incorporates a step-by-step guide through each stage with examples and checklists of the resources and evidence required.

Advance preparation and planning gains time and is vital to ensure that potential threats are identified, assessed and managed effectively. Ideally, the following should be in place: evidence-based protocols and guidance for responding to incidents, protocols for identifying sources of key information for rapid risk assessment, strategies for literature searches and lists of relevant contacts, including named experts.

Rapid risk assessments of potential communicable disease threats can be complex and challenging, as they must be produced within a short time period when information is often limited and circumstances can evolve rapidly. The rapid risk assessment methodology described in this document enables the structured identification of key information using systematic appraisal of the best scientific evidence and/or specialist expert knowledge available at the time in order to provide a clear estimate of the scale of the health risk. This is important in not only communicating the potential magnitude of the risk in a systematic and transparent way, but allows documentation of evidence and gaps in knowledge at the time when the assessment is made.

Reference

European Centre for Disease Prevention and Control. Operational guidance on rapid risk assessment methodology. Stockholm: ECDC; 2011.

Annex 5 – Some challenges of applying the GRADE instrument in a public health setting

The GRADE system has been developed primarily to grade quality of evidence and strength of recommendations concerned with individual-level clinical interventions, and it provides advice on how to move from assessing the quality of evidence, to formulating and grading the strength of recommendations. The system performs very well in this setting. Subsequently, GRADE has been applied to some public health issues like assessment of vaccines and antiviral treatments for influenza. However, broader application of GRADE to the field of public health remains partly untested. When ECDC started exploring the potential use of GRADE beyond the initial fields of GRADE applications to address broader public health issues, like infectious disease risk assessments, causation and spread of infectious diseases, a number of problems have been identified. The table below outlines some of these problems encountered in the process along with proposals on how to solve them. ECDC should continue to work with representatives from the GRADE working group and other public health and evidence-based medicine institutions on further discussing and applying some of the proposed solutions to optimise the assessment of evidence in the broad area of public health.

Problem	Explanation	GRADE comment	Possible solution
The GRADE system is limited to assessing the level of scientific evidence that is only a part of the process of developing a public health guidance.	There are several steps before and after the assessment of scientific evidence that are important in public health decision-making, like prioritisation of topics, selection of experts, dealing with potential conflicts of interests that also should be handled in a systematic, explicit and transparent manner.	Ethical, legal and similar aspects can quite easily be incorporated in the GRADE system as shown with the SIGNs considered judgment form.	ECDC has developed specific procedures to deal with some of those issues. These procedures may need to be further developed as experience of their application accumulates. Further discussions with the GRADE group recommended.
GRADE nomenclature uses terms that may sound pejorative to describe confidence levels of underlying evidence.	Using terms like 'low quality' and 'very low quality' of evidence suggests possible flaws in study design and problems with study conduct. For example the anti-vaccine movement representatives already started using GRADE terminology to convince the public that there is little good quality research supporting e.g. vaccine safety.	It is confusing with different terminology. If the quality of evidence is our degree of confidence in the estimate, we have to convince ourselves with good examples that the quality of evidence really should be upgraded.	More neutral terms could be used e.g. 'level' of evidence. An appropriate terminology is needed to be able to incorporate and value good studies from epidemiology and other fields.
GRADE system seems to have too few categories to capture different levels of quality represented by various study designs beyond randomised controlled trials.	In the current categorisation, a lot of the study designs that are relevant and applicable for public health infectious disease control will inevitably end up in the 'low' or 'very low' quality box.	There is need for better checklists to describe and assess the quality of non-randomised studies, but to ask for a higher resolution at the bottom of the hierarchy will probably not be very helpful.	Instead of asking for more resolution at the bottom of the scale, there is a need to address the issue of levels of quality of evidence differently for different questions. The GRADE approach should be developed laterally to better incorporate other lines of evidence. This issue should be further discussed with the GRADE Working Group.
GRADE system at the moment has no/limited mechanisms to assess the level of evidence of studies other than randomised controlled trials and observational studies.	Some categories of studies cannot be assessed at all by GRADE or will be inevitably deemed 'low' or 'very low quality', e.g. microbiological investigations, health economic models, mathematical models of infectious disease spread, burden of diseases studies, descriptive studies of incidence, prevalence of disease, studies based on surveillance data, including time series analysis, etc.	GRADE also provides advice on diagnosis and resource use. But there are pieces of evidence where we need better guidance.	Other scales, checklists identified as part of this work should be analysed and recommended for different types of evidence. If other tools than GRADE are used in the evidence assessment, a question arises how to pull together the results of these specific assessments. How to transparently weigh the different evidence and information sources should be further explored.

Problem	Explanation	GRADE comment	Possible solution
Limited ways to 'upgrade' quality of observational studies.	Currently there are only limited ways to 'upgrade' the 'quality' of evidence from observational studies, e.g. based on the strength on association, dose response relationship, directness of evidence, lack of possible confounders and consistency of evidence is also being used.	The GRADE working group would welcome good well-documented examples to consider additional rules for upgrading evidence from observational studies and other lines of evidence.	More ways to increase the level of evidence coming from observational studies and other lines of evidence are needed to better account for the variety of these study designs. This can be addressed based on adherence to quality checklists for observational studies and more.
GRADE has limited capacity to be applied to issues beyond the assessment of efficacy and effectiveness (and possibly safety) of interventions.	It is difficult to apply GRADE to assess evidence for the purpose of risk/ threat assessments, causation of disease, infectivity, spread and impact of disease.	Other pieces of evidence could be incorporated into the recommendation process. The challenge is that we have not really identified or developed tools to assess risk or threat assessments, etc. Recommendations focus on interventions or actions, but information on risk, causation, infectivity, etc. may of course influence the formulation and the strength of the recommendation.	Use the tool being developed by the Health Protection Agency and the ECDC EBM Methods Working Group.
Timeliness (time pressure).	GRADE system requires substantial time for full application.	Agree that it is important to develop methods for rapid reviews – rapid guideline processes, when needed.	Develop an 'accelerated/abbreviated grading' scheme for the purpose of rapid risk assessments. The tool being developed by HPA may serve this role.
Scarce evidence (evidence pressure).	GRADE system seems to perform well in situations where there is a reasonable amount of medium- to high-quality evidence. In situations where there is scarce evidence, as very often happens in the field of public health, especially in the epidemiology of infectious diseases and particularly in risk assessment of infectious diseases, it has been difficult to apply GRADE.	If there is insufficient evidence, this must be recognised, and we must make a decision based on what we know. This challenge is not specific to GRADE, but GRADE will help make the problem explicit (identifying lack of evidence for important/critical outcomes – not just reporting what is found in the studies/reviews).	It is important to acknowledge lack of evidence. Decisions often have to be taken in such situations. The principles of GRADE should assist the policymaker to address such situations and to express uncertainties.
The GRADE system has been developed to assist in going from evidence to recommendation, but it does not cover the final stages of the decision-making process.	Going form evidence to recommendation does not only require information about the core scientific evidence but is also based on clinical and public health experience, local context, law, ethics economic and political considerations and more.	GRADE has done quite a lot of work in this field as well, but there is general agreement that it is necessary to improve the methods in this field.	There is a need to further develop tools to make the final stages of the decision-making process better informed, more transparent and explicit.

Annex 6 – Considered judgement forms, SIGN

S I G N	Considered judgement pro forma Strength of Evidence	Part B

Key question:

Outcome measures:

O_1

O_2

O_3

1. Draft recommendation

Draft a recommendation based on the evidence statement (part A, section 7)

Indicate here if there is insufficient evidence (low volume) to make a recommendation and make a recommendation for research in section 6.

2. Judgement on the strength of recommendation

Make a judgement taking into account the factors that can affect the strength of the recommendation.

		O_1	O_2	O_3	O_4	O_5	O_6	O_7	
Quality evidence The higher the quality of evidence, the more likely is a strong recommendation.	High								**strong**
	Mod								
	Low								
	Very low								**weak**

Balance of benefits versus harms and burdens

Explain here the balance of benefits vs harms

The larger the difference between the desirable and undesirable outcomes, the more likely a strong recommendation warranted. The smaller the net benefit and the lower certainty for that benefit, the more likely is a conditional recommendation warranted.	Clearly outweigh	Recommend
	Probably outweigh	Consider
	Not known	**Make a recommendation for research** *(section 6)* **Input to DUETS** *(section 7)*
	Probably don't outweigh	Consider against
	Clearly don't outweigh	Recommend against

Are the net benefits worth the costs

Outline here the costs of the intervention

The higher the costs of an intervention – that is, the more resources consumed the more likely is a conditional recommendation warranted	Yes	Recommend/consider
	Not known	**Involve SHTG/SMC** *(section 8)*
	No	Recommend/consider against

3. Contextual issues

Comment here on the applicability of the recommendation in the NHS in Scotland

	Reference
Prevalence	
Resources required (eg training, equipment)	
Patient issues and preferences	
National policies and initiatives	
SMC advice	
Other	

4. Proposed recommendation

Recommendation	**Strength of recommendation** *Please select level*
	STRONG **CONDITIONAL**

5. Footnote

Explain the judgement on the strength of evidence

6. Recommendation for research

7. Input to DUETS *Comment here on the uncertainty of the effects of treatment*	
8. Question for SHTG/SMC	
9. Final recommendation	
	STRONG **CONDITIONAL**
10. **Footnote** *Explain the nature of the post-consultation revisions*	

Annex 7 – Guideline Evaluation Tool (GET5)

Guideline Evaluation Tool (GET5)

SECTION A – DOCUMENT (GUIDELINES) DETAILS

Guideline title .....

.....

Author .....

Publisher .....

Date of publication M M - Y Y

Type of Document Guideline ☐ Other ☐ (please state).....

Guideline Aim .....

.....

SECTION B – EVALUATING THE DOCUMENT

Reviewer's name .....

Date of evaluation M M - Y Y

Circumstances of this evaluation (please tick)

Scheduled annual review ☐ During an incident/outbreak ☐

	Yes (1)	No/Not Known (0)
1. Is the Evaluation date less than 3 years since the Guideline was published?	☐	☐
2. Does the Guideline meet its aims/objectives?	☐	☐
3. Are the recommendations linked to supporting evidence?	☐	☐
4. Has the Guideline been reviewed by experts prior to publication?	☐	☐
5. Are the recommendations clear and specific?	☐	☐
Total score	☐	

Note: Please use the total score above to grade the guideline accordig to the colour coded below

Guidance Grade - Colour Code (please tick)

GREEN (Total score 4-5)	☐	The document is recommended.
AMBER (Total score 2-3)	☐	The document may be used but only in the absence of better guidelines.
RED (Total score 0-1)	☐	This document is not recommended. Other guideline should be sought.

SECTION C – OVERALL ASSESSMENT

1. What are you looking for from this Guideline?

.....

.....

.....

2. Does the Guideline meet your needs/inquiry? Yes ☐ No ☐

Comments .....

.....

.....

3. Is there anything lacking in the Guideline? Yes ☐ No ☐

If Yes, please explain .....

.....

.....

4. In light of what you know about this topic, should a new Guideline be considered for this subject area?

Yes ☐ No ☐

If Yes, please explain .....

.....

.....

If new Guideline is required, please contact your Guideline Development Group.

Health Protection Scotland

Clifton House, Clifton Place,
Glasgow, G3 7LN
United Kingdom
T. +44 (0) 141 300 1100

Annex 8 – AGREE II PH and GET5

AGREE II PH (Standard and Short) and GET5 [for appraisal AND/OR for development methods]

Domain	AGREE II Item	AGREE II PH (Standard)	AGREE II PH (Short): GET5
Scope and purpose	1. The overall objective(s) of the guideline is (are) specifically described	No change	Essential
	2. The health question(s) covered by the guideline is (are) specifically described	Change to Key questions	
	3. The population (patients, public, etc.) to whom the guideline is meant to apply is specifically described	No change	
Stakeholder involvement	4. The guideline development group includes individuals from all the relevant professional groups	Extend to emphasise transparency in selection	Essential
	5. The views and preferences of the target population (patients, public, etc.) have been sought	Extend to emphasise importance of public views and democratic accountability	
	6. The guideline has been piloted among end users		
Rigour of development	7. Systematic methods were used to search for evidence	Time pressures [see Chapter 2]	
	8. The criteria for selecting the evidence are clearly described	Time and Evidence pressures [see Chapter 2]	
	9. The strengths and limitations of the body of evidence are described	No change [see Chapter 3]	
	10. The methods for formulating the recommendations are clearly described	See Chapter 5	
	11. The health benefits, side effects and risks have been considered in formulating the recommendations	See Chapter 3. See also SIGN considered judgement form	
		Add item re ethics, equity and equality	
	12. There is an explicit link between the recommendations and the supporting evidence	See SIGN and Chapter 5	Essential
		Additional criterion regarding modelling and scenario analysis, and certainty of predictions, etc.	
	13. The guideline has been externally reviewed by experts prior to its publication	No change	
	14. A procedure for updating the guideline is provided	No change	
Clarity of presentation	15. The recommendations are specific and unambiguous	No change	Essential
	16. The different options for management of the condition or health issue are clearly presented	Change to 'intervention'	
	17. Key recommendations are easily identifiable		
Applicability	18. The guideline describes facilitators and barriers to its application	Expand to cover: • Legal and regulatory frameworks • Delivery structures • National policy and politics	
	19. The guidelines provides advice and/or tools on how the recommendations can be put into practice	No change	
	20. The potential resource implications of applying the recommendations have been considered	No change	
	21. The guideline presents monitoring and/ or auditing criteria	No change	
		Add a criterion regarding communication with the public	
Editorial independence	22. The views of the funding body have not influenced the content of the guideline	Extend to emphasise transparency	
	23. Competing interests of guideline development members have been recorded and addressed	No change	